

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 2, 2002, 22:45:52 ; Search time 2977.34 Seconds
(without alignments)
99.736 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaaaagtattccgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_or.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	18	100.0	1604	1	MBU35021	U35021 Mycobacteri
2	18	100.0	453	1	MTU43540	U43540 Mycobacteri
3	18	100.0	9281	1	MHR351	U35017 Mycobacteri
C 4	18	100.0	9764	1	AF006921	AF006921 Mycobacte
C 5	18	100.0	10019	1	AF041819	AF041819 Mycobacte
C 6	18	100.0	12412	6	186264	186264 Sequence 18
C 7	18	100.0	17783	1	AE007028	AE007028 Mycobacte
8	16.4	91.1	1535	6	AK096713	AK096713 Sequence
9	16.4	91.1	1535	6	AK147694	AK147694 Sequence
10	16.4	91.1	16384	1	AF007160	AF007160 Mycobacte
11	16.4	91.1	42096	3	CEP21A3	281594 Caenorhabdi
C 12	16.4	91.1	42437	1	MTY4336	295586 Mycobacteri
C 13	16.4	91.1	48380	1	MTY13E12	274410 Mycobacteri
C 14	16.4	91.1	43401	1	MTY13E12	295586 Mycobacteri
15	16	88.9	4125	9	AK027649	AK027649 Homo sapi
16	16	88.9	4288	9	AK051548	AK051548 Homo sapi
17	16	88.9	4288	9	AK050515	AK050515 Homo sapi
18	16	88.9	184584	2	AL360171	AL360171 Homo sapi
19	15.4	85.6	1689	1	KPN252298	A1292298 Kl-chsteli
20	15.4	85.6	2154	9	B63424	B63424 Homo sapien
21	15.4	85.6	2155	10	KNISK3A	X53427 Rat mRNA 10
22	15.4	85.6	2169	9	HIMG1YSYN	BC004783 Mus muscu
C 23	15.4	85.6	3357	10	HC004783	AF048728 Homo sapi
24	15.4	85.6	26857	9	AK048728	AC010688 Drosophil
25	15.4	85.6	94707	2	AC010688	AC006486 Homo sapi
C 26	15.4	85.6	111084	9	AC006486	AC005401 Homo sapi
C 27	15.4	85.6	113688	9	AC005401	AC008771 Homo sapi
C 28	15.4	85.6	124169	9	AC008771	AF000748 Homo sapi
C 29	15.4	85.6	128077	9	AP000548	AL442110 Oryza sat
C 30	15.4	85.6	140072	8	H0512801	AC069054 Homo sapi
C 31	15.4	85.6	14236	2	AC069054	AC025749 Homo sapi
C 32	15.4	85.6	156861	2	AC025739	AC025087 Homo sapi
C 33	15.4	85.6	160270	33	AC025087	AC090944 Homo sapi
34	15.4	85.6	162507	9	AC090944	AC068283 Homo sapi
35	15.4	85.6	167108	2	AC068283	AL132801 Homo sapi
C 36	15.4	85.6	167760	2	CNS01DTN	AC024168 Homo sapi
C 37	15.4	85.6	189271	9	AC024168	AC073094 Homo sapi
C 38	15.4	85.6	186583	2	AC073094	AC090560 Homo sapi
C 39	15.4	85.6	189015	2	AC090560	AC084432 Homo sapi
C 40	15.4	85.6	190317	2	AC084432	AC036147 Mus muscu
C 41	15.4	85.6	223034	2	AC036147	AC008482 Homo sapi
C 42	15.4	85.6	237231	2	AC008482	AL592185 Mus muscu
C 43	15.4	85.6	258094	2	AL592185	AF004526 Drosophil
44	15.4	85.6	280987	3	AF004526	U49864 Chlamydomon
45	15	84.4	4813	8	CHE019864	

ALIGNMENTS

RESULT 1	MBU35021	1604 bp	DNA	BCT	27-APR-1996
LOCUS	Mycobacterium bovis BCG	DNA	flanking deletion region 3.		
DEFINITION	U35021				
ACCESSION	U35021				
VERSION	U35021.1	GI:1049243			
KEYWORDS					
SOURCE	Mycobacterium bovis BCG.				
ORGANISM	Mycobacterium bovis				
REFERENCE					
AUTHORS	Mahairas,G.C., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.				
TITLE	Molecular analysis of genetic differences between Mycobacterium				
JOURNAL	bovis BCG and virulent M. bovis				
MEDLINE	J. Bacteriol. 178 (5), 1274-1282 (1996)				
REFERENCE	96200095				
AUTHORS	2 (bases 1 to 1604)				
	Mahairas,G.C., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.				

```

TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES   Location/Qualifiers
source     1..1604
            /organism="Mycobacterium bovis"
            /sub_species="BCG"
            /db_xref="taxon:1765"
            /clone="pIS116"
            /gene="rfbA"
            /gene="rmlB"
            /gene="rmlC"
            /note="previously known as rfbB"
            /codon_start=1
            /transl_table=11
            /protein_id="AAB66648.1"
            /db_xref="GI:2326950"
            /translation="MRLIVTGGAGFICTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
            AIVEDAI RLVOGDITDAELVSOLVAESDAVHFAAESHVNDALNDPEPLHTNVIGTFT
            TILAEVRRGVRLLHI STDEVYGDLEDDRAARTETSTPNPSSPYSATKACADMVLVRA
            WYRSCVVRAT1SN:SNNYGYPYUVEKFFIPROITNCLTGLLKLKLYGGGNNVWVHVDH
            NSAVRILDRGRIGETYLISSEGERDNLTVLSLLRLMDRDPDDFDHVTDRVAHDLRY
            AIDPSRIYDELICWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
            /CDS
            150..1139
            /gene="rmlB"
            /codon_start=1
            /transl_table=11
            /product="rhamnose biosynthesis protein"
            /protein_id="AAB66647.1"
            /db_xref="GI:2326949"
            /translation="MKLIIVTGGAGFICTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
            AIVEDAI RLVOGDITDAELVSOLVAESDAVHFAAESHVNDALNDPEPLHTNVIGTFT
            TILAEVRRGVRLLHI STDEVYGDLEDDRAARTETSTPNPSSPYSATKACADMVLVRA
            WYRSCVVRAT1SN:SNNYGYPYUVEKFFIPROITNCLTGLLKLKLYGGGNNVWVHVDH
            NSAVRILDRGRIGETYLISSEGERDNLTVLSLLRLMDRDPDDFDHVTDRVAHDLRY
            AIDPSRIYDELICWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
            /gene="rmlC"
            1136..1747
            /gene="rmlC"
            /note="previously known as rfbC"
            /codon_start=1
            /transl_table=11
            /protein_id="AAB66649.1"
            /db_xref="GI:2326951"
            /translation="MREPRELDVPGAWELTTHVDSRGLFEMLTDHCFKAFAGHS
            LUDRVQNCYSVAGYLRLHFAQLPPOAKYVTCVSGVDFVVDIREGSPYFGRWDS
            VILDDQDRNIYSDGLAHGFLAQDNSTWYLCASRYNPQRQHTICATDPTLASIKRW
            SNGGCPDSISDRDAAPSFEEVRRGLLPGGTDSRFIGEMRGTT"
            /repeat_region
            2202..2655
            /note="3-4 copies in M. tuberculosis genome; Found in the
            rfb cluster (inside the rfbA gene), also found in the mce
            of M. tuberculosis; present at 3-4 sites of the genome of
            H37Kv, H37Ra, Erdman, and many isolates; not found in M.
            smegmatis & M. avium."
            /rpt_type="dispersed"
            /rpt_unit="1806..2250
            558 t
            679 a 1226 c 990 g
            BASE COUNT
            ORIGIN

Query Match      100.0%; Score 18; DH 1; Length 3453;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcaaaagtgtatcgcg 18
    |||||
DB 2216 ACATCAAAAGTGTATTCGG 2233

RESULT 3
MBDR3S1
LOCUS      MBDR3S1 9281 bp DNA BCT 27-APR-1996
DEFINITION Mycobacterium bovis deletion region 3, 5' end.
ACCESSION  U35017
VERSION     U35017.1 GI:1049238
KEYWORDS   .
SEGMENT    1 of 2
SOURCE     Mycobacterium bovis.
ORGANISM   Mycobacterium bovis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE  1 (bases 1 to 9281)
            Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
            Molecular analysis of genetic differences between Mycobacterium
            bovis BCG and virulent M. bovis
            J. Bacteriol. 178 (5), 1274-1282 (1996)
            96200095
            JOURNAL
            MEDLINE
            REFERENCE
            2 (bases 1 to 9281)
            Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
```

```

TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES   Location/Qualifiers
source     1..1604
            /organism="Mycobacterium bovis"
            /sub_species="BCG"
            /db_xref="taxon:1765"
            /clone="pIS116"
            /gene="rfbA"
            /gene="rmlB"
            /gene="rmlC"
            /note="previously known as rfbB"
            /codon_start=1
            /transl_table=11
            /protein_id="AAB66648.1"
            /db_xref="GI:2326950"
            /translation="MRLIVTGGAGFICTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
            AIVEDAI RLVOGDITDAELVSOLVAESDAVHFAAESHVNDALNDPEPLHTNVIGTFT
            TILAEVRRGVRLLHI STDEVYGDLEDDRAARTETSTPNPSSPYSATKACADMVLVRA
            WYRSCVVRAT1SN:SNNYGYPYUVEKFFIPROITNCLTGLLKLKLYGGGNNVWVHVDH
            NSAVRILDRGRIGETYLISSEGERDNLTVLSLLRLMDRDPDDFDHVTDRVAHDLRY
            AIDPSRIYDELICWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
            /CDS
            150..1139
            /gene="rmlB"
            /codon_start=1
            /transl_table=11
            /product="rhamnose biosynthesis protein"
            /protein_id="AAB66647.1"
            /db_xref="GI:2326949"
            /translation="MKLIIVTGGAGFICTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
            AIVEDAI RLVOGDITDAELVSOLVAESDAVHFAAESHVNDALNDPEPLHTNVIGTFT
            TILAEVRRGVRLLHI STDEVYGDLEDDRAARTETSTPNPSSPYSATKACADMVLVRA
            WYRSCVVRAT1SN:SNNYGYPYUVEKFFIPROITNCLTGLLKLKLYGGGNNVWVHVDH
            NSAVRILDRGRIGETYLISSEGERDNLTVLSLLRLMDRDPDDFDHVTDRVAHDLRY
            AIDPSRIYDELICWAPKHTDFERALKRTTIDWYRDNESWLASLKHAGGRYQDAQO"
            /gene="rmlC"
            1136..1747
            /gene="rmlC"
            /note="previously known as rfbC"
            /codon_start=1
            /transl_table=11
            /protein_id="AAB66649.1"
            /db_xref="GI:2326951"
            /translation="MREPRELDVPGAWELTTHVDSRGLFEMLTDHCFKAFAGHS
            LUDRVQNCYSVAGYLRLHFAQLPPOAKYVTCVSGVDFVVDIREGSPYFGRWDS
            VILDDQDRNIYSDGLAHGFLAQDNSTWYLCASRYNPQRQHTICATDPTLASIKRW
            SNGGCPDSISDRDAAPSFEEVRRGLLPGGTDSRFIGEMRGTT"
            /repeat_region
            2202..2655
            /note="3-4 copies in M. tuberculosis genome; Found in the
            rfb cluster (inside the rfbA gene), also found in the mce
            of M. tuberculosis; present at 3-4 sites of the genome of
            H37Kv, H37Ra, Erdman, and many isolates; not found in M.
            smegmatis & M. avium."
            /rpt_type="dispersed"
            /rpt_unit="1806..2250
            558 t
            679 a 1226 c 990 g
            BASE COUNT
            ORIGIN

Query Match      100.0%; Score 18; DB 1; Length 1604;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcaaaagtgtatcgcg 18
    |||||
DB 465 ACATCAAAAGTGTATTCGCG 482

RESULT 2
MTU43540
LOCUS      MTU43540 3453 bp DNA BCT 14-AUG-1997
DEFINITION Mycobacterium tuberculosis rfbA, rhamnose biosynthesis protein
            (rfbA), and rmlC genes, complete cds.
ACCESSION  U43540
VERSION     U43540.1 GI:2326948
KEYWORDS   .
SOURCE     Mycobacterium tuberculosis.
            Mycobacterium tuberculosis.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
            1 (sites)
            Lee,T.
            Rhamnose biosynthetic genes related to a novel repeated sequence of
            Mycobacterium tuberculosis
            Unpublished
            2 (bases 1 to 3453)
            Lee,T.
            Direct Submission
            Submitted (18-DEC-1995) Tae-Yoon Lee, Microbiology, College of
            Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu
            705-035, South Korea
            3 (bases 1 to 3453)
            Lee,T.
            Direct Submission
            Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of
            Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu
            705-035, South Korea
            Sequence update by submitter
            On Aug 14, 1997 this sequence version replaced gi:1304703.
            Location/Qualifiers
            1..3453
            /organism="Mycobacterium tuberculosis"
            /db_xref="taxon:1773"
            /clone="pIS116"
            /clone="pIS116"
            /gene="rfbA"
            /gene="rmlB"
            /gene="rmlC"
            /note="previously known as rfbB"
            /codon_start=1
            /transl_table=11
            /protein_id="AAB66648.1"
            /db_xref="GI:2326950"
            /translation="MRLIVTGGAGFICTGNFVHSAVREHPDDAVTVLDALTYAGRRRESL
```

/product="hypothetical protein"
 /protein_id="AAK44319.1"
 /db_xref="GI:13879143"
 /translation="MSYVKKAPSRRVLRQTESTVYVKGSGLSWRVRFTCTGCLGAVWG
 GREYHAPAGPTGIGTGKIDMTVTVDAGPGRVSRVSEVAANAALFAIVADPKRIIR
 ELJGSGTSCRNLIKPAKLVGVSFKSTKMKFLGDLPTITSKVTALKNEIIVFVSHPIGH
 RWKPFESTSLPTLFRVTEFDYHAAGATKNGKFLKYEMTKSGKNAACATFAIAKAIQSQ
 YAKRA"
 2280. 2873
 /gene="MT0098"
 2280. 2873
 /gene="MT0098"
 /note="similar to Gp:145427; identified by sequence
 similarity; putative"
 /codon_start=1
 /trans_table=11
 /product="methylinthioadenosine nucleosidase, putative"
 /protein_id="AAK44320.1"
 /db_xref="GI:13879144"
 /translation="MPQPNANIHVDALLDAMVPLGTQCVLDVCGGDTLAAALAKRI
 PYVTVAIDAPVLRKRAQKTFANAPTRWELHADIMTAELPNAGFDAAVSNAAALIHITEDTR
 TALSRLGTLTGCTGLAVVTFPTSLRKLGMILTSWACGMANNKVGKWHISAPIRKP
 PHTWLELSHWKALLPGACIRRLYLGRVLVTWRAPV"
 3002. 3772
 /gene="MT0099"
 3002. 3772
 /gene="MT0099"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /trans_table=11
 /product="hypothetical protein"
 /protein_id="AAK44321.1"
 /db_xref="GI:13879145"
 /translation="MAKNONKIRNRWELITGGLGHVTPDDAALAAKALIKASTGLGE
 VWKLRGCGFALGCGSGCAPEDALIMRGKALROAIIIRALGVERIVKAILVIALAAW
 AVWFKFAGCAIQATIDRIIPVLRAGAFKVDQMTVIHALEKALAAKSTMLAIFTGMLA
 AYAVLQVAGVCGIWEILKMKVEYFANVATSTFLPLEVHJLAKGLTTTKVVTFSINVAAV
 VILITSKLFGVGRCKGAKYDVERGHJLLDLERAAMLT"
 3791. 4162
 /gene="MT0099.1"
 3791. 4162
 /gene="MT0099.1"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /trans_table=11
 /product="hypothetical protein"
 /protein_id="AAK44322.1"
 /db_xref="GI:13879146"
 /translation="MKGLKVCVSTSAFALTPARKAKAAKRVNLDQVAKRRRTDEG
 PLTCHQPCSAKALITAAWSROSHVPATHRCPCCAACAAALTCALTAEANSVVE
 VKAPRAATSTLSRAARESK"
 4206. 4973
 /gene="MT0100"
 4206. 4973
 /gene="MT0100"
 /note="similar to Gp:AL009126; identified by sequence
 similarity; putative"
 /codon_start=1
 /trans_table=11
 /product="5-methylthioadenosine
 nucleosidase/S-adenosylhomocysteine nucleosidase,
 putative"
 /protein_id="AAK44323.1"
 /db_xref="GI:13879147"
 /translation="MAVTVGVCAIQDELAYLRGVLDVDAKROOVAQILFDSGLDHPH
 VVIAAAGKVNVTGATLADIAREGCKTIVFTGVAGGLDPELCLIGDIVADRVVCHHID
 HGLDELRJYORCHPTPFIPTERLGYPPVDAIVDRVHKRLDGFPLAIVSTAAAGGGR
 QPRTIYGTGLTGQYIHLCEKTRNRLHLELGHMAVEMGECAVQICASFDTIWIIVRAI
 SDIADAGSDVDENRFVEVAASSARVLLRLPLVLTAC"
 5105. 7390
 /gene="MT0101"
 5105. 7390
 /gene="MT0101"


```
gene
CDS
LLLGODWRTRITELAGLTAGLDTARALPAVTVRVCGATGVIECDRPPVDLAVATPAA
3120..4280
/gene="biof"
/db_xref="GI:2791846"
/EC_number="2.3.1.47"
/note="Biof; similar to MTCY336.34c"
/codon_start=1
/transl_table=11
/product="putative 8-amino-7-oxononanoate synthase"
/protein_id="AAB96957.1"
/db_xref="GI:2791846"
/transl_table="MKAATQARIDDSPLAWDVAQORHQREAGLRKLRPRPAVATELD
LASNDYGLSRHPAIVDGGVQALRWGAGATGSLRTVTDTKLHQQFAELAEFFGAAV
GLFSSGYTNLNAVGLSGPGSLVSDARSHASLVDPACHLSRAKVVVTPRHVDVDAAD
AALKSREORAVVYDVSFNSDGLAPVRELLFVCRKHGALLVDAANGILGVRCGGHG
LXYELGAGAPDVVMTTTLKALGSGGVGLGPPVPRKHLIDRAKPIIFDTGLAPAAV
GAARALRLQAEPQAVQVNLNHAGELARMCVAAVPDSAMVSIIGEPESAVAAAAA
CLDAGVVGCFRPPTVPAGTSRLRLTARASINAGELELARRVLTDLAVARR"
4277..4957
/gene="biof"
4277..4957
/gene="biof"
/EC_number="6.3.3.3"
/note="Biof; similar to MTCY336.33c"
/codon_start=1
/transl_table=11
/product="putative dethiobiotin synthetase"
/protein_id="AAB96958.1"
/db_xref="GI:2791847"
/transl_table="MTILVVTGTGTGKTVVCAALASAAKAGQIDVAVCKPVQGTGA
RGDDDLAEVGLAGVLTQGLAGLYPOPMAFAAAAHACMALPARIQIVRLIADLDKRG
RLTLVEGAGLLVELAPGVTLRDVAVDAAVALVVVTADLGTLNHPKLTLEALAAQ
VSCAGLVGSGWPDPPGLVAASNRSALARLATVRAALPAGAALDAGDFAAMSAAAFDR
NWAAGLVG"
4957..5466
/note="ORF6; similar to MTCY336.32c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96959.1"
/db_xref="GI:2791848"
/transl_table="MWISIELVFNSDTEAALIRRIWAGLILAAALHPSQAVASRPHYSIAV
AERTAPEDEPLGAVARKLPIUDCVIGAPVIGRANVVTPLVVPTESELLALHAKVHRL
CGPHLAPAPMANSLPGOWTAHVTLARRVGGHQLGKALKIACGRPSKILHRPAGLRWNG
NTRAEYLLG"
complement(5612..6478)
/note="ORF7; similar to MTCY251.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96960.1"
/db_xref="GI:2791849"
/transl_table="MDWLHPDCDLTUTERARKRGITLSNOOYDGMSKI.SGYLTTPQARA
TFEAVLAKAAPGATNPDDHTPVIDTTPDAAAIDRDTKSAQANIDGILAGLRALIAS
GELGONGLPVSIUVTTLTDLQTCAGCKGFTGCGTLLPMADVIHMTSHAHYSPASGR
YQOALFDGHTPLALHTRKLASPAQRIMLFANDRGCTKPGCUAPAYSHAHYVGTWIS
TGRTDITDLTLACDPDNRKLAEGWTRKNTHGHTENLWPPPHLDHCQPPKINTFPHHEKL
LRHNDEDNDDP"
complement(6339..7007)
/note="ORF8; similar to MTCY336.16"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96961.1"
/db_xref="GI:2791850"
/transl_table="MLANSREELVEVFDALDAELDRLDEVSFFVLTTPERLRSLERL
CLVRLPAVGHITLNLDPQOASEEELGCTLCCALANRLRTFKPDAALKJADAADLGP
RALTCPELAPOLATATAQROGLICEAHIKVIIRALFRPPARKGCVHPHGRSRPGRO
SRSSSRAGRLRPAGGLATPRKPRHRTKPTORHHHPQPAIRKKHVTAKWLPDPPS
AgHL"
7455..8504
gene
```

```

CDS
/gene="bioB"
7455..8504
/note="BioB; similar to MTCY336.15c"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAB96962.1"
/db_xref="GI:2791851"
/transl_table="MTQAAVUKPTNDAGQDGNNSDILVAAKOOVLQKRGELNOTDVLV
VLQPHDRIPELLALAHFVYKMKCGPIVEVEGILSLKTCGCPHCHFCUSGIFASPV
KSAWLDIPSILVEAKQATKANSKATEPILVAVRGPDDEKLMAQVANGIEAIRKEVEINIA
CSLGMILTAEODVQLAAKGVYKHNINIJETARSEFFANVVVTHITWEERWOTLSMVRDAGME
VYCGGTLGMEETILQURAEFAELAEILGPDEVLNPLNPREGTTIFADILEVMPGDAKKA
VAARILALIPKLMILRFACGRITLCLDCLCAKRGILGGINAVIVGNVLTITIGRPAAEADLEL
LHLOMPLKALNASL"
8825..9406
/note="ORF10; similar to MTCY336.13c"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB96963.1"
/db_xref="GI:2791852"
/transl_table="MGLSATGVLYGCLWAWIAPPIDHAVVAITRACERVHYLYGSESO
NPIIAPMLIGLILSVIAYVASALMKWQREHRRGQMVAGLSIGLTTAAAIAGVGALVV
KIKYGALDFDTVPLSGDHIALTYTQAPPVFPFRRPQLOIALTLMPAGTASLVTALLA
ACTIARDLGGYPAVDPSNNAKTEALETTPQAPVS"
complement(9571..>10019)
/note="ORF11; similar to MTCY336.12"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAB96964.1"
/db_xref="GI:2791853"
/transl_table="RMAGROMGDFLHPELEDILSTPEVSHVFGDTKIGSAVPTPPVLI
VVAHYDYLIVSDIDALADSYTAGGANVYHRDLFSEHVSILHPLSAPMTLKWLTDRFA
GKPLTOHRVPTMTILNPMYAGMARLAVIAAKVITGRKLSKKPL"
BASE COUNT 1476 a 3154 c 3655 g 1774 t
ORIGIN

Query Match 100.0%; Score 18; dB 1; Length 10019;
Best Local Similarity 100.0%; Prod. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acatcaagatgattcagc 18
|||||
bb 6619 ACATCAAGATGATTGCG 6602

RESULT 6
LOCUS 186264 12412 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 18 from patent US 5700683.
ACCESSION 186264
VERSION 186264.1 GI:3205982
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12412)
AUTHORS Stover,C.Kendall and Mahairas,G.G.
TITLE Virulence-attenuating genetic deletions deleted from mycobacterium
BCG
JOURNAL Patent: US 5700683-A 18 DEC-1997;
FEATURES location/Qualifiers
source 1..12412
/organism="unknown"
BASE COUNT 2146 a 4274 c 3875 g 1917 t 200 others
ORIGIN

Query Match 100.0%; Score 18; dB 6; Length 12412;
```

Best Local Similarity 100.08; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 acatcaagtgattgcg 18
|||||
Db 468 ACATCAAGTGATTCGCG 485

RESULT 7
AE007028/c

LOCUS M -cobacterium tuberculosis CDC1551, section 114 of 280 of the complete genome BCT 27-APR-2001
DEFINITION AE007028 AE000516
VERSION AE007028.1 GI:13881250
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 17783)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
Location/Qualifiers
1..17783
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
complement(104..826)
/gene="MT1622.1"
complement(104..826)
/gene="MT1622.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45890.1"
/db_xref="GI:13881251"

translation="MLAKLAPGATNPDDHTPVIDTTPDAAIIDKDKSQQRNHDGL
LAGLRALTASGLGQHNGLPVSIVVTTTLDTLOTGAGGTGGTLLPMADVIRMTSH
AHHYSPASGRVPOALFDHGTPLAYHTRKLSAPQRIMLFANDRCCTKPCGDAPYHS
QAHVHTWTSIGRTDITLTLACDPDNRLAEKGWTHKTHGHTWLPPLPHLDHGQPR
TNTHHKLRLHNDENHDHP"
complement(831..1499)
/gene="MT1623"
complement(831..1499)
/gene="MT1623"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45891.1"
/db_xref="GI:13881252"

translation="MLANSRELVEVDALDAELDRLEDSFEVITTHPERLSRLERLE
CLVRRLPVGHTLNLQDTSASEEIGTTCACLANRIITKPDAAIHKIADAADLGPRL
RALTGEPLAPOLTATATAQROGLICGAHIKIRALRIPPARRKGCVIIIIPGRKSRPGRO

gene
CDS

SKS1SSRRAGPLRPAGHGLATPRRRPHRRTPTQTRHHPEQPAIRRHVTAKWLPDPPS
AGHL"
1947..2996
/gene="MT1624"
1947..2996
/gene="MT1624"
/note="similar to GP:1680650; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAK45892.1"
/db_xref="GI:13881253"

translation="MTQAATRPNTIAGODGNNSDILVYVAKQVQLQKQEGINQDOVLA
VIQLPDDHLEELIALAHIEVRMRWCPEVEGIIISLTKGGCPEDCHFCQSGLFASPV
RSWLIDTISLVEAAKQATSGATEFCIVAARGPDPERLMAQVAAGLIEAIRNEVETNIA
CSLGMITAEQVLDAAARGVIRYNNHLEIARSFFANVTTHWEERWOTISXVYDAGME
VCCGGIICMGCTIQAQAEFAELAFIPGDEPLNPLNRPGTTFADIKVMIPVGDALKL
VAARFIALPRTMLKFPAGGREITLGDLSGAKRGILGGINAVIVUNYIITPLGRVAEADLEI
LDELQMLKALNASI"
2997..3236
/gene="MT1625"
2997..3236
/gene="MT1625"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45893.1"
/db_xref="GI:13881254"

translation="MVEIVAGKQRAIWAAGVYVNYVTGELADTATPTAARMLLEPRRFC
AAQGRMVVGVKPIQGWAKCSHHGQVDSADIATQR"
3317..3898
/gene="MT1626"
3317..3898
/gene="MT1626"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45894.1"
/db_xref="GI:13881255"

translation="MIATISATGVLVGLWMIATPPIHAVVAITRACERVHHYKGSQ
NFEIAPFMALGLASVLAUVSALMWREHRCPOMVAGLSIGLTTAAAIACGVCALVV
KLYGALIDFVLSKSHQIALTYVIGAPVFEPARPLQIALIMVAVGLASLVYVALLA
AIALADHDAITTAVIDPSSNAKTEALHTQAIVS"
complement(3953..4054)
/gene="MT1627"
complement(3953..4054)
/gene="MT1627"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45895.1"
/db_xref="GI:13881256"

translation="MRSATGTANSDLAHLNLYTPDLNQFDLPLESKR"
complement(4063..5430)
/gene="MT1628"
complement(4063..5430)
/gene="MT1628"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45896.1"
/db_xref="GI:13881257"

translation="MPDGEGLSLMVEPGLAGATGAEWIGRPPHEELQKVRPLPSPD
DPFYPPAGYQHAVPGTVLRSRDVELAFMGLIPOPVTATOLLYRTTNMGNPEATVTT
VIVAEALPGQTCPLLSYQCADANSSRCFFSYALRRBAKALGSLTOWELIMISAALA
EGWAVSPDGPKGLGSPYEPGVRLVDGIRALNSRVSGLSPATPGLIWKYSGGGL
ASAWAAACGYPADILDTVGAVGSPVLDGHTFRRLNGTLIAGLPAIVAAIQHSYIP
GLIARYIKEHANDEGROIIEQLTEMTYDVAIVRMAGRDMGDFIIDEPLDIISTPYVSHV

FGDTKLGSVPTPPVLIQVAHVHYLDIVSDIDALADSVTAGGANVTYHRLDFSEHVSIL
HPLSAPMTLRLWLTDRFAGKPLTDHVRVTWPTIFNPMYAGMARLAVIAAKVITGRKL
SRRL"

gene complement(5660. .6193)

CDS /gene="MT1629"

/gene="MT1629"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK45897.1"

/db_xref="GI:13881258"

/translation="MTSSVRRLAEKVDLRELALHLEQLAVFSDPHRLPGIRMIASVYL

GVVSPATPELPADRWHPVSSLPMAFDGPMVTHARTLIAKMSVNTICFALAPKE

FALSTLRDIYGAALGYQVDATNLQRLVLRVITQTCTIAQSGSGCKHPAALRYFTDS

QLRVTFDEFAALRPPGQL"

6419. .7468

gene="/gene="MT1630"

CDS /gene="MT1630"

6419. .7468

/note="similar to SP:P11458 GB:X12713 PID:581137 GB:U00096

PID:1651334; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="quinolinate synthetase"

/protein_id="AAK45898.1"

/db_xref="GI:13881259"

/translation="MTVLNRTDTLVDELTAADITNTPLGCGVGDDGDEHWAETIRLAHL

RGATVLAANYQLPAIQDVADRVGDSLSRVAAPEDTIVFCGVHMAETAKILSPH

KTVLPDQAGCSLADSIPTDELRAWKDEHFGVAVVSYVNTAAVKALTDICCTSSNA

VDVVASIDPDEVLFPCDPQFLGAHVRYTGRKNLHVWAGECIVHAGTNGDELAQARA

HPDAEFVHPGECATSALYLAGEGAFPAERVKILSTGCMLEFAAHTTRAKOVLVATEV

GMLHOLRAAPEVDPRAYNDRASCYKMKWTPAALLKCLIVEGAEVHDVPGIAASGR

SVQRMIEICHPGGE"

7468. .9051

gene="/gene="MT1631"

CDS /gene="MT1631"

7468. .9051

/note="similar to GP:608530; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=11

/product="L-aspartate oxidase"

/protein_id="AAK45899.1"

/db_xref="GI:13881260"

/translation="MAGPAWRDAADVVVIGTCVAGLAALAAADRAGRSVVVLSKAAQT

HYTATHYAGGGLAVVLDPNDDSDVAHVADTLAAGAGLCPDAVYSIVADGYRATVDLV

CAGARLDSVPGRWALTREGHSRRRIHVAGDGTATCAEVORALODPAAGMLDIRTCHVA

LRVLHDGTAFTGLLVVRPDGCGIISAPSVILATGGLGHLIYATNPNAGSTGDTALGL

WAGVAVSDLEFIQFHTMLFAGRAGRRPLITEAIRGEGAILVDROGNSITAGVHPMG

DLAPRDVVAALIDARKLKATGDPVYLDARGIEGFASRFTPTVASCRAAGIDPVRQ?IP

VVPGAHYSCGGIVTDVYGOTELLGLYAAAGEVARTLGHANRLASNSLI.EGLVVGGRAG

KAAAHAAAGRSRATSWPEPISYTLDRGLQRAMSRDASHYRAAAGLHRLCDS

LSGAQVRDVACRRDFEDVALTLVAQSVTAALARTESRGCHHRAEYPTCTVPEQARSIV

VRGADANAVQVALVVC"

9021. .9908

gene="/gene="MT1632"

CDS /gene="MT1632"

9021. .9908

/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262

PID:456041; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

Query Match 100.0%; Score 18; DB 1; Length 17783;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18

|||||

DB 1111 ACATCAAGTGATTCGCG 1094

RESULT 8

AR096713

LOCUS AR096713 1535 bp DNA

DEFINITION Sequence 1 from patent US 6008201.

AR096713

ACCESSION AR096713.1

VERSION AR096713.1

KEYWORDS GI:10025749

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

1 (bases 1 to 1535)

AUTHORS Riley,L.W.

TITLE DNA molecule encoding for cellular uptake of mycobacterium

tuberculosis and uses thereof

JOURNAL Patent: US 6008201-A 1 28-DEC-1999;

FEATURES location/Qualifiers

source 1..1535

BASE COUNT 297 a 544 c 458 g 236 t

ORIGIN /organism="unknown"

Query Match 91.1%; Score 16.4; DB 6; Length 1535;

Best Local Similarity 94.4%; Pred. No. 95;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18

|||||

DB 841 ACGTCAAGTGATTCGCG 858

RESULT 9

AR147694

LOCUS AR147694 1535 bp DNA

DEFINITION Sequence 1 from patent US 6224881.

AR147694

ACCESSION AR147694.1

VERSION AR147694.1

KEYWORDS GI:15111784

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

1 (bases 1 to 1535)

AUTHORS Riley,L.W. and Chong,P.

TITLE DNA molecule fragments encoding for cellular uptake of

Mycobacterium tuberculosis and uses thereof

JOURNAL Patent: US 6224881-A 1 01-MAY-2001;

FEATURES location/Qualifiers

source 1..1535

BASE COUNT 297 a 544 c 458 g 236 t

ORIGIN /organism="unknown"

Query Match 91.1%; Score 16.4; DB 6; Length 1535;

Best Local Similarity 94.4%; Pred. No. 95;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaatgattgcgcg 18

|||||

DB 841 ACGTCAAGTGATTCGCG 858

RESULT 10

AE007160

LOCUS AE007160 16384 bp DNA

DEFINITION Mycobacterium tuberculosis CDC1551, section 246 of the

complete genome.

ACCESSION AE007160 AE000516

VERSION AE007160.1

KEYWORDS GI:13883389

SOURCE Mycobacterium tuberculosis CDC1551.

ORGANISM Mycobacterium tuberculosis CDC1551.

Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS

1 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.

Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains

JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.

Direct Submission

TITLE
JOURNAL

Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source

1..16384 Location/Qualifiers
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"

gene

96..839
/gene="MT3557"

CDS

96..839
/gene="MT3557"

/note="similar to SP:Q00298; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="serine esterase, cutinase family"

/protein_id="AAK47897.1"

/db_xref="GI:13883390"

/translation="MGAGALITAVLLIAGVMTIVAFADGCPDAIEVTFARGTEPPP
GIRGVQAFVDSLRQQTGMEIGYIPVNYAASRLQLHCGDCANDAI SHIKSMASCCPNT
KLVLGYSOGATVIDIVAGVPLGCSIFSGPLPAAYADNVAAVAFVCFNSNRAGGSLS
LSPFLGSKAIDLCNPTDPICVHGCPGNEFSGHIDDIPTVTYTTQAAASFVQRLRAGSVPH
LPGSVPLPGSVLQMPGTAAPAPESLHR"

gene

871..1566
/gene="MT3559"

CDS

871..1566
/gene="MT3559"

/note="similar to SP:Q00298; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="serine esterase, cutinase family"

/protein_id="AAK47898.1"

/db_xref="GI:13883391"

/translation="MRFICVIPPPOHSGRWAGARRLTSILVAAFAAATLLTPAL
APPASGCDPAEVFARGTGERPGLRGVGAOFVSSLROOTNKSICTGYVNYPANGDEL
AADGANDSDHIQQNASACRATRLVGGTSGQAAVIDIVTAAPILPGIGFTQPLPAA
DHIAAIALFGNPSGRAGGLSALTPOFGSKTINLCNGDPCSDGNRRRAH/GYVPG
MTNQAARFVASRI"

gene

1624..1863
/gene="MT3560"

CDS

1624..1863
/gene="MT3560"

/note="identified by Glimmer2; putative"

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK47899.1"

/db_xref="GI:13883392"

/translation="MTAADWHTAYHCLMLAPDAPQARRPSTQPNAHLSAECAAMNRN
HFTGREVCGHANRTGLTSTAALRRVDAGRHNKQ"

gene

1835..3523
/gene="MT3561"

CDS

1835..3523
/gene="MT3561"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK47900.1"

/db_xref="GI:13883393"

/translation="MMPGVJTNSESPTAADHDIRTATRETLEDYTLRLAPKSYRRWPP
AVGSLAGGIAYIADFAIGANVGITWGTANALGCIATFALVVPVGLPIAYAAAYN
IDDLITRGSGFGYGVSVTVNIFATFTIFFALEGSIIMAGCIKILILHPIIAGVACYS
TILITPLVYVGMKVISQIOWMTPLMLMAAPFGYIVSHPDSTGQFPFVYAGKDHG
GILSGSVLLAAGVILSLIAIAEQIDYLRNPPRTPENAKRWMTTILLAGVWAFGA
"KQIITGLFVLAIVLMANIPGSSITANQPVIOFMQIYTFVPGMLALTILAVILVLSQIK
INVTNAYSCSIANTNSFTRIKHYIGRVVFLGNLAIALILMEANMPDFNTILCCYA
NCGMAVVAASDIFGNKYLGLSPKTFEPRGMLYAINPVCGSILAAAGLSIVTFE
CGICAAIQPSPLVALVVALVMPPIILAAATKGKYYLRTHDCIDLPWYDEHGNPSAAV
LTCVCHODFERPDMACQTHAHVCSLCLSTDKQAEHVLPLGAHAIIPGDOVP"

gene

/gene="MT3562"

CDS

/gene="MT3562"

/note="MT3562"

/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U00096
PID:1788657; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="LkNA pseudouridine synthase A"

/protein_id="AAK47901.1"

/db_xref="GI:13883394"

/translation="MSLTKRPKSPORPRISGVVRLRLDIAYDGTDFACMAAQVCO
KIVAGDIDAALTITPTPVNLRAGRTDAGVHAGQVAHVDPADALPNAYPRAGHVG
DYEPFLRLRLRIGLGFIPADVRIIDITTRAPAGEDARFSALRRIITVYRLSLATPYGVEPQA
HFTLAMPRELIDIDAMTAASHDLMGCIIDFAFCRHREGATTIRDIQRLJMSRAGATLTA
IIVTADAFQWSMVRSILVAGVGHRRATTWCRELLTATGRSSDFAVAFHGLTILQV
IYPPDIQILASNNLVTBQVRSK"

gene

/gene="MT3563"

CDS

/gene="MT3563"

/note="MT3563"

/note="similar to GB:J01685 SP:P02416 GB:X00766 PID:147716
PID:42800; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="ribosomal protein L17"

/protein_id="AAK47902.1"

/db_xref="GI:14883395"

/translation="HUKPIKGPRIAGSSSHKQKLIANLATSILFHCRTITTPPKARAL
IKYAKLLITHAKKALINKREVLKKLRDQVVIITLFAEIGPFAIDKACGYTKLIKIEA
IKGHNAPMAVLELVKEKIVTSEANKARVAAQAQAKAKKAAAMPTESEAKIVAEKGDV
CASEPDAAKAPERPTPEAPN"

gene

/gene="MT3564"

CDS

/gene="MT3564"

/note="MT3564"

/note="similar to SP:P72404; identified by sequence
similarity; putative"

/codon_start=1

/transl_table=11

/product="DNA-directed RNA polymerase, alpha subunit"

/protein_id="AAK47903.1"

/db_xref="GI:13883396"

/translation="MLISORPTLSEDLTDNRSFVIEPLEPGFCYTLGNSLRLTLLS
SIPGAAYTSIRIDGVHEFTTTPVPGVKEDVTEILLNLKSLVVSSEIEHIVTMYLKQGP
GHVTAGDIPVPAGVTVHNPHHATLNDKRLLEVELVVERGHGYVAVQNRAGAEIG
RIPVDSIYSPVILKVTYKVDATRVEQRTDFKLILDVETKNSISPNDALASCHAKTIVEL
FGLARELNVEAGIEIGSPAEDHIAFALPIDDLDLTVRSYNSICKREGVHTVGEV
AYDEODYAEIOL"

gene

/gene="MT3565"

CDS

/gene="MT3565"

/note="MT3565"

/note="similar to SP:P81288; identified by sequence

REMARK	The C.elegans Sequencing Consortium. Erratum:[published errata appear in Science 1999 Jan 1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep 3;285(5433):14931] 2 (bases 1 to 32096) McMurray, A.A. Direct Submission Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1HQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rws@nematode.wustl.edu On Jul 24, 1997 this sequence version replaced gi:1695020. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: this sequence is not the entire insert of clone F21A3. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone F21A3 is at 1 in this sequence. The true right end of clone F21A3 is at 6434 in this sequence. The true left end of clone F28F8 is at 31993 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z81463. The end of this sequence (31993..32096) overlaps with the start of sequence Z81071. For a graphical representation of this sequence and its analysis see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F21A3 .
FEATURES	Location/Qualifiers 1..32096 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="v" /clone="F21A3" complement(216..288) /gene="F21A3.1,2" /note="CAC Val V RNA predicted using tRNAscan-SE-1.11 preliminary prediction similar to tRNA Val" complement(216..288) /gene="F21A3.1,2" 211..2178 /note="F21A3.1,1" /note="CAT Met M-tRNA predicted using tRNAscan-SE-1.11 preliminary prediction similar to tRNA-Met" 211..2178 /gene="F21A3.1,1" join(2426..2501,2545..2664,2708..2842,2972..3056, 3109..3201,3355..3506,3863..4074) /gene="F21A3.3" join(2426..2501,2545..2664,2708..2842,2972..3056, 3109..3201,3355..3506,3863..4074) /gene="F21A3.3" /note="similar to EGF-domain protein" /codon_start=1 /protein_id="CAF04157.1" /db_xref="GI:3876159" /db_xref="SPTREMBL:O17829" /translation="WTRINIVFLSLAVCTYSOTWTKTIDFCSPSPFNSSCGPGCK MEISLGNRCCEPFLGMRKQPCQDVYKSCARSEERCHWTRPISLFFADWALISG GCQKNGKQALALIP ILDRIDEMFVGVRSEKTSAAHRFEPMSGPKYKFIIDVGI SFDQPIV SVAKPTIIVGHTDVHVEFLTKPFDHETGVFLNKPKDIDILVS ILVT NTDGLMILEGCTVKGITQ IRLQETVYKGMAGVFRDEIVKSRMFNL IANSL IERRVAVM
SOURCE	
tRNA	
gene	
tRNA	
gene	
CDS	

```

similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAK47904.1"
/db_xref="GI:13883397"
/translation="MARTGPVTWKKSRLLTDLVGGIOAFEXKPYPHJQHRARIKES
EYLLOQRKQARETYGMVEKQPKFYEEAVRQPCKTGTEILKLIESKRIDNVIYRAGL
ARTRMARQLSHGHFNHGVPNSVSQYDVIIVIROKSLNTVPFQIAKETAGER
PIPSVLQVGSGRRRLVLIHQLPERRAQIDVPLTEQLIVEYYSK"
complement(6706..7125)
/gene="WT3566"
complement(6706..7125)
/gene="WT3566"
/note="similar to SP:p45812; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S11"
/protein_id="AAK47905.1"
/db_xref="GI:13883398"
/translation="MPPAKKGPNATSAARKGKTTRREKKNPVIGAAHI KSTFNNTI VTI
TPQGCVITAWASSGHGVFKSRSKSTPFAAQLAAENAKRAKDHCVRKKVDVFVKGPCGS
RTETARSQAAGLEVGAISDVTPQHNGVRPPKKRV"
complement(7129..7503)
/gene="WT3567"
complement(7129..7503)
/gene="WT3567"
/note="similar to GP:2104383; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S13"
/protein_id="AAK47906.1"
/db_xref="GI:13883399"
/translation="MARLVGVLDLRPRDKRMEVALITYINCIGTKPSNEIIAATGIDRDILA
TDRLDEQGLIHRLROYIEANLKAVEGLDRREVQADIRRK IECVCYGGLHKRGMPVRQOR
TKTNARTRKGPRTIAGKKKAR"
complement(7719..7832)
/gene="WT3567.1"
complement(7719..7832)
/gene="WT3567.1"
/note="similar to SP:p45810; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11

Query Match          91.1%; Score 16.4; DB 1; Length 16484;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Zy      1 acatcaagtgattcgcg 18
Db      11346 ACGTCACAGTCATCCGC 11363

RESULT 11
CEFP21A3       CEF21A3     32096 bp    DNA        INV      25-OCT-2000
LOCUS         Caenorhabditis elegans cosmid F21A3, complete sequence.
DEFINITION
ACCESSION Z81509
VERSION   Z81509.1 GI:2276160
KEYWORDS HTG.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
            Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
            1 (bases 1 to 32096)
REFERENCE
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for
        investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
```

```
GRGVTTKKKRTFV"
join(14078, .14402,14665, .15074,15212, .15475)
/genes="F21A3.1"
CDS
join(14078, .14402,14665, .15074,15212, .15475)
/genes="F21A3.1"
/notes="predicted using GeneFinder
contains similarity to Pfam domain: PF01604 (7TM
chemoreceptor), Score=254.2, E-value=5.6e-73, N=1"
/codon_start=1
/protein_id="CAB04154.1"
/db_xref="GI:3876156"
/db_xref="SPTREMBL:O17826"
/translation="MCMNTSFAKFLDTQFLSTTMYTTLIICLPVHLIFGCLGVFKTP
TOMCNKMPMLTLLWSADLSDFGLVFPFVQPVLAGYSGILNGHGVPLRVWYIPD
CVQIQVIGVCALEFETRFILYAKOSCKHLRRPRLTLNTICVYMPIYSLIPD
KQTKGAVQFSKCLQPOEYDEKVELLTWTSTGYAYNMSLNTAVOQTLIFVLIYNNM
KSNVKEVSKKRLDMRDTFLKTIILQVTLPLTVNGPLMLNFFTINYSYNOGANNL
SISMMATHGLVSLAMLYLHKSWATLLOILSPKFIREDRETMMI:SKVHOHQOFS"
complement(join(15549, .15743,15934, .16071,16247, .16444,
16540, .16598,16681, .16823,16995, .17181,17233, .17415,
17532, .17703))
/genes="F21A3.2"
complement(join(15549, .15743,15934, .16071,16247, .16444,
16540, .16598,16681, .16823,16995, .17181,17233, .17415,
17532, .17703))
/genes="F21A3.2"
/notes="predicted using GeneFinder
similarity to Aspergillus acid phosphatase (PK:Q12546)
cDNA EST yk394.5 comes from this gene
cDNA EST yk69f10.5 comes from this gene
cDNA EST yk231b3.5 comes from this gene
cDNA EST yk285e4.5 comes from this gene
cDNA EST yk539e5.5 comes from this gene"
/codon_start=1
/protein_id="CAB04153.1"
/db_xref="GI:3876155"
/db_xref="SPTREMBL:O17825"
/translation="MTYDDTLKSIYEGTDSLHSVGRCAVFLDGKHSVMRYIH
RVNLTGLPGRFYHVGSDHGSPIFFETALKEREDGGLYAVYGDIGVENGSRSLGH
TOKWAQGLDMVLHVGDFAYNMDESNGETGGEFFRQIPIRVAGYIIMYATVGNHYYN
NFTHYVARETMNSHNLFSYDQVPVHFVFEFTFFYFQNGTHOMENQYMLINDL
KVLRLLDQSOQLSQSPQMDHNGHGSTHIRTGLPLTHGYALEKILFYFYGVDEIWA
HEHSVRLPVNYRTVNGTRHPYDPPAPVHIITGSAGCRENTDVFVHPPPWSAVR
STDYGFQVMRYNSTHLNFKOINVAOECKLDDDFVWVWSSPKHRI:FKIKRLRTY
GTHIPKIHCHHSKCHDCTRKRKRHSRSHVEI"
join(21375, .21425,21480, .21710,22206, .22358,22411, .22695)
/genes="F21A3.4"
join(21375, .21425,21480, .21710,22206, .22358,22411, .22695)
/genes="F21A3.4"
/notes="predicted using GeneFinder
Similarity to Human dystrophin (SW:P11532)"
/codon_start=1
/protein_id="CAB04155.1"
/db_xref="GI:3876157"
/db_xref="SPTREMBL:O17827"
/translation="MDIKVIREKI:TLTTSITSESVAFYEHVDPCCKLLEHLELEME
KKIOLSRIDVVQSVQELWDRVDKARRRSDISCRVTAWAQLHSENMNVLKKAKRM
CTTWEISATLAEIQESLIDLAERKVIQVRMEIDYRRHLHQQLKDLKLDGDRREE
KCYWNDRKSALIEKVEMLKQSKLSGERVEERKMEQKKMLEEMLLKKTKGRR
QLIDENANLVSFSLTEESS"
join(22967, .23146,23217, .23582,24576, .24916,26030, .26153,
26204, .26320,27040, .27510)
/genes="F21A3.5"
join(22967, .23146,23217, .23582,24576, .24916,26030, .26153,
26204, .26320,27040, .27510)
/genes="F21A3.5"
/notes="predicted using GeneFinder
weak similarity to protein kinases"
/codon_start=1
/protein_id="CAB04156.1"
/db_xref="GI:3876158"
/db_xref="SPTREMBL:O17828"
/translation="MNRNRKGDLENIEHRTDDEIAGLKPKQLFNGKWLILDEKKEKR
```

```
GAACVYLVCIDKPKCKKPGILYLEICEQNVTTIANQVDYFHQOOSLQYSHRFSALIDAGI
INNHVFMVVRIRAGTITLHULLKCLSSDKMSVTTASFLAVIMISALIELLSASGWLNRN
FISKQMLDITKTOFYLAADATDITVSSDKKHHRAIDELHLRTAFSLHLHWKJGDIJYAP
KSFVYDKDSHMTDITVJMMEMMLYVITDWTGICKLPWKSXSKSKKELIMKKELFLPNLQKE
PEFTNKVYEQQIDVDVMDITIALRNFPAKILKVAKEEQEKLEKLPVRGGAWFKGPKAGNQ
ISTVNYHGTIIDDYFKIVCSGRPAWALHWRDVMIDWDRKLIJNTIHFETSKMFEAYPKHQKS
LEISEMERIQATKEHYTMKNHITFTHMAKNQAAIVEYILMPFEAEPEFTDKKKDPEE
FAAAAVGKKRGKKPKKKIDDPKMKELDKVLDKDFVVGICIQVTEFSKSKLHMSKRRF
PTNCRKNWNCRFAYAKLIANGCI"
complement(join(28770, .28389,28450, .28539,28591, .28708,
29458, .29458))
/genes="F21A3.6"
complement(join(28770, .28389,28450, .28539,28591, .28708,
29458, .29458))
/genes="F21A3.6"
/notes="predicted using GeneFinder
Weak similarity to Shark myoglobin (SW:P02206), contains
similarity to Pfam domain: PF00042 (Globin), Score=39.5,
E-value=1.5e-10, N=1"
/codon_start=1
/protein_id="CAB04152.1"
/db_xref="GI:4876154"
/db_xref="SPTREMBL:O17824"
/translation="MYRQILHQHIEHVNVQDARPONSLEKPTILIEHKEUKSKSEKPV
TILVFKPAKVAEKIKPTTSSEKENAVPDPLARKIITDTSRI:SHQHWLVGKTAFLIQL
ICVINGIKLFWRLFSYPRYKLIWQFRAFLDSSMLNAVALRHHSVYLVGKJACKLIDS
MKDEALCKSMKSLAVAHIKWNQVRNHTVSDRRTYFEQGFHVGAKRPVAVICSGRLRP
KVASLSAEMRGRTGRLAEALMPKRIPEALINGCLMLKQTSIDNLLKFTATFQIMIEPVL
KVVRENGYQLDHTKQAWVLYGVYIADLIEVFCRALND"
join(complement(2024, .2053),complement(1600, .1709),
complement(441, .478),complement(48, .196),
complement(281463,1:38584, .38668),
complement(281463,1:38131, .38424),
complement(281463,1:37206, .37251),
complement(281463,1:36933, .37109),
complement(281463,1:36634, .36807))
/genes="F21A3.7"
/notes="Weak similarity to N-terminus of acetylcholine
receptor proteins, contains similarity to Pfam domain:
```

Query Match 91.1%; Score 16.4; DB 3; Length 32096;
Best Local Similarity 94.4%; Pfam No. 1.5e-02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acetylcholine receptor 18
DB 21379 AFATCAAGATGATTCGCG 21379

RESID: 12
MTCY336/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

MTCY336 32437 bp DNA BCT 03-AUG-2001
Mycobacterium tuberculosis H37Rv complete genome: segment 70/162.
295586 AL121456
295586.1 GI:3261785
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 32437)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,K., Devlin,K., Felwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogi,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skellton,S.,
Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S., and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Nature. 393 (6685), 537-544 (1998)
98295987
2 (bases 1 to 32437)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2117233.

COMMENT

NC's:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source
Location/Qualifiers
1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..273
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="v48"
33..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="v336"
complement(39..2228)
/gene="Rv1565c"
complement(39..2228)
/gene="Rv1565c"
/note="Rv1565c" (MTCY336.48), len: 729. Unknown membrane protein, some similarity to 005402 HYPTHEPICAL 72.2 KD PROTEIN from B. subtilis (634 aa) opt: 184 E(): 4.8e-17; 29.1% identity in 378 aa overlap and to hypothetical protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, X392_HAEIN P43993 h10392 (245 aa) opt: 265; E(): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse score is 0.930
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1565c"
/protein_id="CAB09072.1"
/db_xref="GI:2117272"
/db_xref="SPTREMBL:006625"
/translation="MLTSPRPALTPPEPALPVMVTGRTTGTGYRHDLDGLNGVAIALVAVFVWFGVSGVDVFLASGFFFGKILRAALNPDLISPIAEVIRLIRRLPLA LVVSLVAGLITTAIOPQTWEAFANOSLASGYQNWELASTVNLRAGEAVSPLO HNSVQGOFLYLAIFLLVAGCAYLLRLRPAPRYLRFTVLLSTLTLASFIYAI VARIHQATAYNTFARAWELLAGALGVAVVPRWPMRLRTAVATAALAILSCCAL IDYKPEPGPVALVPWLTMLILAGANROCHGCTRDRLPLPNRLLATAPLVALGAMA YSWYLMWPLLIIFWLSYTGRRHANFVEGAVALVLSGLLITLRLVEDPLRLRAPAGY SRDYPGARALIDHVRVPLKRMPTVLEVRDLPTSTKDCISDFVNPALINCTYGDV APRTIALAGGASHWLTALDLCRRHHFKVTVLKMGCPLSTEEVPLMGNPNVAPQ CHWQVQAMAKVADHPDYVFTSTRPNWIKPGDVMPATVYGLWQTFADNNIPFLAMR DTPMLVLDQGPFLPADCLAGGNPQSCGIARSKVLIDKNIPLDFVAHFLPLKPLDMSD AICRTDTCRAVEGNVLVYRSHHLTPYIMRTMTSELGKQIAANTIMW"

gene
CDS

complement(2327..3019)
/gene="Rv1566c"
complement(2327..3019)
/gene="Rv1566c"
/note="Rv1566c" (MTCY336.37), len: 230. Function: unknown probably exported has QOAV repeats at C-terminus, similar to hypothetical M. leprae protein, to putative invasins 1.2(007390, 00791) from M. avium and slightly similar to C-terminus of listeria invasion-associated protein p60 precursor p60_LISMO P21171. FASTA results, Q49634 COSMID B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa overlap. TBparse score is 0.890"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAH09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:006624"
/translation="MKSKMSQSFAICLAMLAPMVAAPCLAAADPATPRVDYQOITD VILAGLSQGVLPKSNAGGILSGPTHGTGTGINVGFADSGILQYAYACAGILPRSS GMYKKQKVLPOAKKGDILFYGPFGTQSVALYLGKGMLEFQDVGVVSPVPTNCMT PYLVRVIGTQPTPVQVAPVQVAPVQVAPVQVAPVQVAPVQVAPVQVAPVQVAP VQPPPTAKSR"
complement(4040..4046)
/note="possible RHS, AAGGAGG, for Rv1566c"
complement(3258..3263)
/note="ASNI Site: APTAAT; probably linking fragments R1/G8"
complement(3259..3543)
/note="Rv1567c"
complement(3259..3543)
/gene="Rv1567c"
/note="Rv1567c" (MTCY336.36), len: 94. Unknown. Thpase score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:006623"
/translation="MVTMTSPSELFAFTNVCPPDACPLVPFGVNYIYPPVMGGTG AATAVIGHPVSMIKQWYNSFWPIISIAVITVTSIAGYAIACFSEHYWH"
1791..5104
/gene="bioA"
1791..5104
/gene="bioA"
/note="bioA" (MTCY336.35c), bioA, len: 477. Function: probable adenine-thymine-8 amino-7-oxonucleotide amino transferase (B2.6.1.62) 006622. Contains aminotransferases class-III pyridoxal-phosphate attachment site (PS00600). FASTA results, BIOA_MYCLE P4548 (436 aa) opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also similar to other M. tuberculosis proteins e.g. MTCY227.12c (449 aa). FASTA score: E():3.5e-16; 29.5% identity in 421 aa overlap. TBparse score is 0.874"
/codon_start=1
/transl_table=11
/product="bioA"
/protein_id="CAB09069.1"
/db_xref="GI:2117269"
/db_xref="SWISS-PROT:006622"
/translation="MAATGSLTPEQIITAVDGAHLWHPYSSIGREAVSPVVAHGA WITLIDGQPIEVLDMSSWTAIHGHGHPALDQALTLQLRVMHVFEGGITHPAAR LAKLVDITPAGLDVTFSSDGSVSVEAAKMAQWYRGRLPGKRRRLMTWRGYYHCD TFLMSICDGHGHSLSLTLAAQVFAFPVPRDYPSAFAEQALQAHLGALAAV VEPVQVAGGRFHPDPRYLHDLRCRREYVLLIFDEATIGFRTGLFAADHAGVSP DLMCVKALTGSLAALCTADVAHTISAGAALMHGPTFMANILACAVASVSE ILLGDWTRITELAGLCTAGLTARALPANTVDRVCAIGVIFCDRVDILAVATPAA IDRGVLRPFRNLVYAMPYICTPREITQIISAMVEVARLVCLP"
4541..4654
/gene="bioA"
/note="PS00600 Aminotransferases class-III pyridoxal-phosphate attachment site"

misc_feature


```
/gene="hycQ"
1629..3095
CDS
/note="Rv0086, (MTCY251.04), oxidoreductase, len: 488,
hycQ, most similar to sp|P12999|BIOC_ECOLI BIOTIN SYNTHESIS
COMPONENT F(BC 1.- (526 aa), FASTA scores, opt: 948
z-score: 1117.4 E(): 0; 35.9% identity in 493 aa overlap.
Also similar to E. coli d9087711 & NUOL_ECOLI P33607 nadh
dehydrogenase chain 1 (613 aa) FASTA scores, opt: 360,
z-score: 354.9 E(): 3.2e-13, (27.9% identity in 488 aa
overlap), and to NUON_ECOLI P33608, nadh dehydrogenase i
chain n (425 aa), FASTAScores, opt: 375, z-score: 371.4,
E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/translation="MTGLLLAAIAPLAASIASLTGMRRTATLTALSATTVLACAV
AMGFWMGSAQFGLGLRADALTVMVLVIGVGTLTATAASIGVITELARHGHIDGR
SARLVGLTAPFLCAMLAVCANNGIVWIAETVITAFVGHRRTRTALEATWKY
VVICSGVIAVRLGTVLVFAARDSCAAAGALNDLILAEHAAGLDPCGVARLGGLLL
IGGARAGLPFHTLWADAHQAAPAPVASALMSGVLAVAFSVLIRLRPIILDVAGSPAY
LRNGLLVAGLTLLAVLMLTVTGDKVRLMAYSSMEHMLGLIAAAGCTLTALAALL
HVLAHGIGTKTVLFGAGQQAHDSTADITGVWRRSRRLGVSHAVGLIVLGLPPF
AMPASGLATARSILANERLAWLGAALLLIATIGTALTARNRMLLGTFAACAPATVP
ATAAALMWGIVVSAALGITAGPLADLLGIAASNVGLP"
3092..4570
gene
/ gene="hycE"
3092..4570
CDS
/ gene="hycE"
/ note="Rv0087, (MTCY251.05), oxidoreductase, len:
492, hycE, similar to HYCE_ECOLI P16431 formate
hydrogenylase subunit 5 (569 aa) FASTA scores, opt: 680,
z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa
overlap) and to NUOD_ECOLI P33600 nadh dehydrogenase 1
chain d, (407 aa) FASTA scores, opt: 245, z-score: 293.1,
E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/translation="MMSASVLRHVRERGLIATAEQIWAHSFRLAIWAHDDSDSLRV
VTLRAGLPDRVEVYVVDNPEIRSLAYLSFPAGRFEREMADLYGIRPVGHKPR
RLVRHAWPDWHPKRTDAGPAPEFTDGAFFLAVECPGVYFIPVGVHAGIIEPCHF
RFSVAGETIVRLKARLWVVRHGEKLFHGPRATAVADIAKRIISCDTSAAHALASHLAI
EDALGIELPHEVRLRALI VELEKLYNHAADLGCALANDVGYSLANAHOPIRENLRR
NAAVTGHRLLRAGTRAGVALRALPI TDELAALADLAEVATLTILANSVYVDFRAGTA
VLRPDDASALCGLGYVARASGLSDARVEHTI VLPITEIGAPDGVILARYVVRDEF
AASALAQHVESHTGPIEVAATLHPVGPAPSSGIGVEGWRGTIVIRHVEIDVDGRITR
AKVDPSPKHNKPALPVAMADTIVPDPFLANKSFNSQSVAGNDL"
4605..5279
gene
/ gene="Rv0088"
4605..5279
CDS
/ gene="Rv0088"
/ note="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/translation="MSVYKAPSRVRLQRSTVVKVGRSSLSRRVRVTTGDLGLVWG
GREGYRVPKPTGPQKGMVTVVDVADGPRSVRSVEVAAPAEALFAIVADPPRRH
ELDGSRTVGNIKVPKLVGSKFSTKMKFLGPLYRTSKLKNELVNSPLHGLH
RWRWFESLSPILTRVTETFDYHAAGAINKGLKTYEMTGFAKSNAAAGTEATLAKLSQ
YARGRA"
5436..6029
gene
/ gene="Rv0089"
5436..6029
CDS
```

```
/gene="Rv0089"
/ note="Rv0089, (MTCY251.07), len: 197. unknown, some
similarity to sp|P12999|BIOC_ECOLI BIOTIN SYNTHESIS
PROTEIN BIOC. (251 aa). FASTA scores: opt: 202 z-score:
262.0 E(): 4.5e-07; 39.0% identity in 118 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0089"
/protein_id="CAA98925.1"
/db_xref="GI:1405759"
/translation="MIPWPNANIHYDALDAMVPLGTCVLDVGGCDGLLAARLARRI
IVTAVDIDAPVLRKRAQIFRANAPI RNLHADIMTAEPLNAGDAVVSNAALJHHIEDTR
TALSRLGGLVTPGCTILAVVTEVTPSIRNLGLMHLTSWVACGMANRVKGMESAPIKWP
PQCTLHELBSHVRALLPGACIRRLVGLVLTWRAPV"
6147..6152
RBS
/ note="aaggaa, potential rbs upstream of Rv0090"
6158..6928
gene
/ gene="Rv0090"
6158..6928
CDS
/ gene="Rv0090"
/ note="Rv0090, (MTCY251.08), len: 256. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0090"
Query Match 91.1% Score 16.4; DB 1; Length 38380;
Best Local Similarity 94.4%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 acatacaatgatacgcg 18
DB 12392 ACGTCAAATGATTCGGG 12375
RESULT 14
LOCUS MTY13E12 43401 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
ACCESSION Z95390 AL123456
VERSION Z95390.1 GI:3261766
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
AUTHORS Cole, S. T., Brosch, K., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S. V., Eigmeier, K., Gas, S., Barry III, C. E.,
Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Mout, S., Murphy, L., Oliver, S., Osborne, J., Ouali, M. A.,
Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skellton, S.,
Squares, S., Squares, R., Sulston, J. E., Taylor, K., Whitehead, S. and
Barrell, B. G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)
JOURNAL 98295987
MEDLINE 2 (bases 1 to 43401)
REFERENCE Parkhill, J.
AUTHORS Direct Submission
TITLE Submitted (11-JUN-1998) submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
JOURNAL On Jun 27, 1998 this sequence version replaced gi:2104370.
COMMENT
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide web.
```

(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBpares (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source

Location/Qualifiers

1. .43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"

source

<1. .1481
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y77"

gene

78. .1481
/gene="Rv3448"

misc_feature

78. .164
/gene="Rv3448"
/note="PS00402 Binding-protein-dependent transport systems:inner membrane comp signature"

CDS

78. .1481
/gene="Rv3448"
/note="Rv3448, (MTCY77.20), len: 467 aa. Unknown membrane protein, contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature. Some similarity to AL021930|MTV035_18 Mycobacterium tuberculosis (472 aa) fasta score, opt: 429 z-score: 488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3448"
/protein_id="CAB08736.1"
/db_xref="GI:3261767"
/db_xref="SPTREMBL:O33354"
/translation="MPTSDPGLRRVTIACAGQAVH.TLPAAVPIVATLIPSLVDILACR
GASPTAARYOLSGAPALPNATLAUCGLRDCGAVLVLIKSSACIPTPRCDVAAAV
AAADLTARPOCQRTRELSGALASCTAGGGLMLVKNVAGTNVTRYSDIATAGVAAA
GLAALLFAVTCRTROPFAGLTLSVITATIFGAVAGILAVPGVGVSHSLVAAQMAAA
TSVLAMRTGGGTLTAVACCAVVAATLIVGATITAPVPAIGSLATLASPGLLEVS
ARMAYLLAGLSPLRPALNPDDADALPTDRLITRANKRADWLTSLIAAFASATITGA
IGTAVATGCIHRSSMGGTALAAVTGALLLRARSADTTRRSIVFAICGTTTATATTA
ADRALEHGPWTAAITAMLAAMVFLGFVAPALSLSPVTYRTIELLECIATAMVPLTA
WLCGAYSARVHLDLTWT"

source

611. .>43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y13E12"
1478. .2845
/gene="Rv3449"
1478. .2845
/gene="Rv3449"

gene

1478. .2845
/note="Rv3449, (MTCY13E12.02), len: 455. Function:
probable precursor of serine protease. Has putative signal
peptide N-terminus and hydrophobic stretch at
C-terminus. Contains three signatures typical of subtilase
family: aspartic acid active site (PS00136), histidine
active site (PS00137), serine active site (PS00138). FASTA
results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E():
8e-07; (38.0% identity in 387 aa overlap), similar to
MTCY15F10.29 (45.7% identity in 451 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3449"

CDS

/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/translation="MTTSKTRLLRVVSAALATLSGLGTPVAHANVSPPIIDERWLPESAL
IAPRPPTVQREVCTEVTAEASGRAFSQAQLADLDVWRLTRGAGUKVAVIDTGA
KHRLPKVAVAGDYPFTGDTGADCAHKTAAAGTAAAGPDAQSNFSCVAPWTLTISI
KQSSKPAFVGDIPSTGVDYDTMAKAVKTAADLCASVTNISSIACTVFAAAPDDRAL
CAALAYADVKNVAVAAAGNTGGAACPPQAPGVTHDSVTVAVS/PAWYIDYVITVGS
VNAUCPSAFTLACHVWVAVANTGEAVTSLSPFGDGTNNRLGQRGSITPISCTSYAAPV
VSGAALIRAKRPPIITAKVVMORTESAHIPPAWHPPLVGVGNTVDIALAANVSSISLPOA
CTATSDPAFVAVLVPRHNSPPGPSDKRALHTAFAGAACLLALMATIATASAKRLRPRGN
GIAGI)"
1757. .1789
/gene="Rv3449"
/note="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862. .1894
/gene="Rv3449"
/note="PS00137 Serine proteases, subtilase family,
histidine active site"
2456. .2488
/gene="Rv3449"
/note="PS00138 Serine proteases, subtilase family, serine
active site"
/complement(2810. .4222)
/gene="Rv3450c"
/complement(2810. .4222)
/gene="Rv3450c"
/note="Rv3450c, (MTCY13E12.03c), len: 470. Function:
unknown but similar to Rv3895c MTCY15F10.17 (FASTA score:
1.6e-27; 36.4% identity in 475 aa overlap); possible
membrane spanning region near N-terminus, and to
Y14967|MLH8628_16 Mycobacterium leprae cosmid H628; (481
aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 42.9%
identity in 480 aa overlap, also similar to Rv3869
(MTCY027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:O06317"
/translation="MPSVATWIVHNSGYHFLRRIRKALLFGVCAATGALRARTTSL
ALGVLATVAARAKATVALLRGSALGQAPIVMGRESGALYVYVDIVWHVVLINLASAR
LIAATHANIPVPSSEGLHHRKGLTIGTGAPQLLIGPLAGAESAMACDSNGSGSTT
VVGFAFSSSAVLAEGRIHVAATESGPTLYLGGRAVVIJAHIVVMALRQGRV
THVVAUSLLNAVPEAPKTPAPRIKRGGRASVGLPGFLVGGVYVRIITKASDEYVVLIED
GVQRIGQVAADLLRPVUSGCVSNVITVAPDVIKVPVIVNTLPVSAFPPKPTPTVDSGSP
GHAVTTLCTVTPAQCACARAVFLAGSGPPVPLGGVPVTLAQADIGKHALDAVYLPFG
KSAVAAKSLISGCGTCTRYLVTDTCVRFATHDDVAHDLGLPTAAIPIAPWPLATLPS
GPELSKANAASVARDTVAPGP"
/complement(4049. .4081)
/gene="Rv3450c"
/note="PS00139 Prokaryotic membrane lipoprotein lipid
attachment site"
/complement(4228. .4231)
/note="Possible RBS, GGAG, for Rv3450c"
4388. .5131
/gene="Rv3451"
4388. .5131
/gene="Rv3451"
/note="Rv3451, (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUN2_MYCTU_Q50664
probable cutinase cy339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:49:09 ; Search time 366.19 Seconds
(without alignments)
42.142 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaagtgattcgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_1101.*
1:	/SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	12412	17	BCG deletion regio
2	16.4	91.1	1535	16	Mycobacterium tube
3	16.4	91.1	1335	17	M. tuberculosis ce
4	16.4	91.1	1535	19	DNA for M. tubercu
5	16	88.9	2642	22	Human EST-derived
6	16	88.9	2753	21	Human ORFX ORF2359
7	16	88.9	3125	22	Human cDNA sequenc
8	16	88.9	3262	21	Human secreted pro
9	15.4	85.6	10594	20	Enterococcus faeca
10	14.8	82.2	639	18	H. pylori flagella
11	14.8	82.2	1731	18	H. pylori flagella

12	14.8	82.2	96109	22	AAF28548	Genomic fragment #
13	14.4	80.0	21	22	AAF98936	Human gene single
c 14	14.4	80.0	69	21	AAA48232	T. reesei xylanase
c 15	14.4	80.0	573	19	AAV36098	DNA sequence of th
c 16	14.4	80.0	596	21	AAA48219	T. reesei xylanase
c 17	14.4	80.0	624	22	AAH56886	P. patens lipid met
c 18	14.4	80.0	625	22	AAH50894	Lipid biosynthesis
c 19	14.4	80.0	787	21	AAAC35047	Arabidopsis thalia
c 20	14.4	80.0	1516	21	AAAC33367	Arabidopsis thalia
c 21	14.4	80.0	1669	21	AAAC39469	Arabidopsis thalia
c 22	14.4	80.0	1682	21	AAAC35565	Arabidopsis thalia
c 23	14.4	80.0	2361	19	AAV70952	DNA sequence of th
c 24	14	77.8	300	20	AAZ12931	Human gene express
c 25	14	77.8	491	21	AAAB9724	Human ovarian carc
c 26	14	77.8	545	22	AAF93587	Lung carcinoma cDN
c 27	14	77.8	841	22	AAH03727	Human cDNA clone (
c 28	14	77.8	853	22	AAH07512	Human cDNA clone (
c 29	14	77.8	1753	21	AAAC43745	Zea mays DNA fragm
c 30	14	77.8	1959	21	AAH07582	MariGold beta-cycl
c 31	14	77.8	2152	21	AAAC74845	Human ORFX ORF400
c 32	14	77.8	2667	22	AAH16538	Human cDNA sequenc
c 33	14	77.8	2890	22	AAH82205	Melon ethylene rec
c 34	14	77.8	3797	22	AAH82217	Melon ethylene rec
c 35	14	77.8	4140	10	AAH92735	Sheep PrP gene for
c 36	14	77.8	11126	18	AAV74482	Staphylococcus aur
c 37	13.8	76.7	32	20	AAH87640	PCR primer lrc375-
c 38	13.8	76.7	32	22	AAF72869	Primer #8. Homo s
c 39	13.8	76.7	390	11	AAO05314	3' portion of sequ
c 40	13.8	76.7	401	21	AAAC30364	Human secreted pro
c 41	13.8	76.7	642	22	AAH02160	Shigella flexneri
c 42	13.8	76.7	730	21	AAAC34255	Arabidopsis thalia
c 43	13.8	76.7	806	22	AAH07001	Human cDNA clone (
c 44	13.8	76.7	828	22	AAH04867	Human cDNA clone (
c 45	13.8	76.7	866	21	AAA54169	Sequence encoding

ALIGNMENTS

RESULT 1	
AAAT33537	
ID	AAAT33537 standard; DNA; 12412 BP.
XX	
AC	AAAT33537;
XX	
DT	15-FEB-1998 (first entry)
XX	
DE	BCG deletion region 3 and flanking sequences.
XX	
KW	BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW	mycobacteria; vaccine; infection; marker; ss.
XX	
OS	Mycobacterium bovis strain BCG.
XX	
PH	Key Location/Qualifiers
FT	misc_feature 1406..10673
FT	/tag=
FT	/note= "BCG delta 1 deletion region"
XX	
PN	W09625519-Al.
XX	
PD	22-AUG-1996.
XX	
PF	15-FEB-1996; 96WO-US01938.
XX	
PR	17-FEB-1995; 95US-0390878.
XX	
PA	(PATH-) PATHOGENESIS CORP.
XX	
PI	Mahairas GG, Stover CK;
XX	
DR	WPI; 1996-393419/39.
XX	

PT Detecting markers for avirulence in Mycobacterium - used in
PT production of vaccines against bacterial infection, and to detect
XX bacterial infection
XX
PS Example 1; Fig 3; 66pp; English.
XX
CC This DNA sequence comprises Mycobacterium bovis BCG deletion
CC sequence BCGdelta3. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 Other
CC deletion regions (see AAT33535 and AAT33536) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absence of deletion junctions (see AAT33538-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
XX
SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;

Query Match 100.0%; Score 18; DB 17; Length 12412;
Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0;

Qy 1 acatcaaaagtgttcgcg 18
Db 468 acatcaaaagtgttcgcg 485

RESULT 2
AAQ89200
ID AAQ89200 standard; DNA; 1535 BP.
XX
AC AAQ89200;
XX
DT 09-OCT-1995 (first entry)
XX
DE Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
DE protein.
XX
KW Vaccine; tuberculosis; ss.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 1..1535
FT /*tag= a
XX
PN WO9506726-A.
XX
PD 09-MAR-1995.
XX
PF 01-SEP-1994; 94WO-US09863.
XX
PR 02-SEP-1993; 93US-0118442.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Riley LW;
XX
XX WPI; 1995-115442/15.
DR P-PSDB; AAR71931.
XX
XX DNA encoding for cellular uptake of Mycobacterium tuberculosis -
PT used to develop prods for vaccines, passive immunisation and
PT diagnosis and cellular uptake of other materials
XX
PS Claim 2; Page 9-11; 46pp; English.

XX The isolated DNA molecule of the invention confers on M.
CC tuberculosis an ability to enter cells and to survive within
CC macrophages. It encodes a polypeptide having a mol. wt. of about 50-
CC 55 kDa, pref. 52 kDa. The AA sequence represents a highly
CC hydrophilic protein with a hydrophobic region at its carboxy
CC terminus. It could be a secreted protein, a cytoplasmic protein, or
CC a surface protein with its carboxy terminus attached to the outer
CC membrane of the organism. The deduced AA sequence is in AAR71931.
XX
SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 16; Length 1535;
Best Local Similarity 94.4%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgttcgcg 18
Db 841 acgtcaaaagtgttcgcg 858

RESULT 3
AAT33656
ID AAT33656 standard; DNA; 1535 BP.
XX
AC AAT33656;
XX
DT 22-NOV-1996 (first entry)
XX
DE M. tuberculosis cellular uptake gene region.
XX
KW Cellular uptake; cell entry; macrophage; passive immunisation;
KW vaccine; gene therapy; ds.
XX
OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
XX
FH Key Location/Qualifiers
FT CDS 181..810
FT /*tag= a
FT /label= ORF-1
FT /product= cell entry protein
FT CDS 886..1535
FT /*tag= b
FT /label= ORF-2
FT /product= macrophage survival protein
XX
PN W09626275-A1.
XX
PD 29-AUG-1996.
XX
PF 20-FEB-1996; 96WO-US02155.
XX
PR 22-FEB-1995; 95US-0392210.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Riley LW;
XX
XX WPI; 1996-425086/42.
DR P-PSDB; AAW02301.
XX
XX DNA giving M. tuberculosis ability to enter mammalian cells -
PT and/or survive within macrophage(s), useful in vaccines to protect
PT mammals against Mycobacterium tuberculosis infection
XX
XX Claim 2; Page 41; 67pp; English.

XX A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an
CC ability to enter mammalian cells and to survive within macrophages.
CC The encoded protein sequence is given in AAW02301. The DNA was obtd.
CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II
CC vector and screening recombinant E. coli strains for Hela cell-

CC invasive clones. The DNA includes 2 separate coding regions (see
 CC also AAT34657-58) coding for the cell entry (AAW02302) and macrophage
 CC survival (AAW02303) proteins. It can be used to produce the cellular
 CC uptake proteins used as vaccines or to facilitate uptake of other
 CC materials, e.g. therapeutic genes.

XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 17; Length 1535;

Best Local Similarity 94.4%; Pred. No. 12;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acatcaaaagtattcgcg 18

|| |||||

DB 841 acgtcaaaagtattcgcg 858

RESULT 4

AAV18647

ID AAV18647 standard; DNA; 1535 BP.

XX AC AAV18647;

DT 03-JUL-1998 (first entry)

DE DNA for M. tuberculosis cellular uptake protein fragment.

XX Cellular uptake protein; vaccine; infection; ds.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT CDS 1..1534

FT /*tag= a

FT /note= "stop codon not given"

XX W09805784-A1.

XX 12-FEB-1998.

XX 06-AUG-1997; 97WO-US13056.

XX 07-AUG-1996; 96US-0689411.

XX (CONN-) CONNAUGHT LAB LTD.

XX (CORR) CORNELL RES FOUND INC.

XX Chong P, Riley LW;

XX WPI; 1998-145620/13.

XX P-PSDB; AAW47541.

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
 XX therapeutic agents e.g. antibiotics, also useful in vaccines

XX Disclosure; Pages 9-10; 82pp; English.

XX The present sequence encodes a Mycobacterium tuberculosis cellular
 XX uptake protein fragment, which confers on M. tuberculosis an
 XX ability to enter mammalian cells and to survive within macrophages.
 XX The protein can be used in a vaccine to prevent M. tuberculosis
 XX infection, and provide for the uptake in cells of, e.g.
 XX antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
 XX raised against it can be used to treat mammals already exposed to
 XX M. tuberculosis, to induce a passive immunity and prevent disease
 XX occurrence.

XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 91.1%; Score 16.4; DB 19; Length 1535;

Best Local Similarity 94.4%; Pred. No. 12;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 5

AAH99048

ID AAH99048 standard; cDNA; 2642 BP.

XX AC AAH99048;

XX 12-OCT-2001 (first entry)

XX Human EST-derived coding sequence SEQ ID NO: 905.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 XX diagnostics; forensic test; gene mapping; genetic disorder;
 XX biodiversity; gene therapy; nutrition; ss.

XX Homo sapiens.

XX W0200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02687.

XX 25-JAN-2000; 2000US-0491404.

XX 17-JUL-2000; 2000US-0617746.

XX 03-AUG-2000; 2000US-0631451.

XX 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

XX P-PSDB; AAM24389.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 XX antibodies and research use -

XX Claim 1; Page 724-724; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
 XX proteins from a variety of organisms, including human, dog, cat, horse,
 XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
 XX from the organism of interest. They can be used in diagnostics,
 XX forensics, gene mapping, identification of mutations, to assess
 XX biodiversity and for nutritional purposes. The present sequence is a cDNA
 XX of the invention.

XX Sequence 2642 BP; 638 A; 703 C; 725 G; 576 T; 0 other;

Query Match 88.9%; Score 16; DB 22; Length 2642;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catcaaaagtattcgcg 17

|| |||||

DB 642 catcaaaagtattcgcg 657

RESULT 6

AAC76804

ID AAC76804 standard; cDNA; 2753 BP.

XX

AC AAC76804;
XX
DT 08-FEB-2001 (first entry)
DE
DE Human ORFX ORF2359 polynucleotide sequence SEQ ID NO:4717.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihyroid;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
DR P-PSDB; AAB42595.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 3901-3903; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB44397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihyroid; and antinaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 2753 BP; 652 A; 746 C; 735 G; 619 T; 1 other;
SQ

Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 catcaaaagtgaac 17
Db 453 catcaaaatgaltcac 468
RESULT 7
AAH17859
ID AAH17859 standard; cDNA; 3125 BP.
XX AAH17859;
AC AAH17859;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17563.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
XX EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
XX Claim 8; SEQ ID 17564; 2547pp; 17b KUM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dp primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 3125 BP; 760 A; 797 C; 848 G; 720 T; 0 other;
SQ

```

Query Match      88.9%; Score 16; DB 22; Length 3125;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaagtgatgcg 17
   |||||
DB 856 catcaagtgatgcg 871

RESULT 8
AAAI16623
ID AAAI16623 standard; cDNA; 3262 BP.
AC AAAI16623;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone pt332_1 nucleotide sequence SEQ ID NO:11.
XX
XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antirheumatic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX
XX WC200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark HF, Fichtel K;
XX
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94903.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing,
PT hematopoiesis regulating, tissue growth, activin/inhibin
PT antiinflammatory or tumor inhibition activity -
XX
XX Claim 20; Page 474-475; 641pp; English.
XX
XX AAAI16618 to AAAI16697 encode the human secreted proteins given in
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans
CC

```

```

and animals. The polynucleotides can be used as markers for tissues in
which the protein is preferentially expressed, as molecular weight
markers on Southern gels, and as chromosome markers or tags to identify
chromosomes or to map gene positions. The proteins can be used in the
treatment of immune deficiencies and disorders, such as severe combined
immunodeficiency (SCID), as well as viral, bacterial, fungal and other
infections. These infectious include human immunodeficiency virus (HIV),
hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
candidiasis. The proteins can be used to treat autoimmune disorders such
as connective tissue disease, multiple sclerosis, systemic lupus
erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
autoimmune inflammatory eye disease. The proteins can also be used to
treat allergic conditions, such as asthma. AAAI16698 to AAAI6774 represent
probes for the human secreted proteins from the present invention.
XX
XX Sequence 3262 BP; 794 A; 859 C; 894 G; 715 T; 0 other;

Query Match      88.9%; Score 16; DB 21; Length 3262;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaagtgatgcg 17
   |||||
DB 978 catcaagtgatgcg 993

RESULT 9
AAAI13377
ID AAAI13377 standard; tNA; 10594 BP.
XX
XX AC AAAI13377;
XX
XX 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO:440.
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
XX
XX W000000000, A2.
XX
XX 12-MAY-1998.
XX
XX 04-MAY-1998; 98WO-US08998.
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044041.
PR 16-MAY-1997; 97US-0046655.
XX
XX (HUMAN ) HUMAN GENOME SCI INC.
XX
XX Harash S, Hilton PJ, Kunsch CA;
XX
XX WPI: 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1667-1672; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAAI2938 to AAAI3919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC

```

CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.

XX Sequence 10594 BP; 3197 A; 1733 C; 2323 G; 3337 T; 4 other;

Query Match 85.6%; Score 15.4; DB 20; Length 10594;
 Best Local Similarity 94.1%; Pred. No. 57;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaatgattcgc 17

|||||

Db 7892 acatcaaaagtgttcgc 7908

RESULT 10

AAT67421
 ID AAT67421 standard; DNA; 639 BP.

XX AC AAT67421;

XX 08-JUL-1997 (first entry)

DE H. pylori flagella-associated protein ORF 19557055.aa.

XX Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW flagellum; ds.

XX Helicobacter pylori.

FH Key Location/Qualifiers

FT CDS 1..639

FT /*tag= a

FT /transl_except= (pos:505..507, aa:Xaa)

FT /transl_except= (pos:514..516, aa:Xaa)

FT /transl_except= (pos:523..525, aa:Xaa)

FT /transl_except= (pos:526..528, aa:Xaa)

FT /transl_except= (pos:535..537, aa:Xaa)

FT /transl_except= (pos:544..546, aa:Xaa)

FT /transl_except= (pos:550..552, aa:Xaa)

FT /note= "Xaa = unknown; no stop codon given"

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

XX P-PSDB: AAW20191.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

XX Claim 1; Pages 163-164; 1481pp; English.

XX The present sequence encodes a Helicobacter pylori flagella-

CC associated protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

XX Sequence 639 BP; 222 A; 122 C; 142 G; 154 T; 9 other;

Query Match 82.2%; Score 14.8; DB 18; Length 639;
 Best Local Similarity 88.9%; Pred. No. 82;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 acatcaaaatgattcgc 18

|||||

Db 110 acatcaaaactcattcgc 127

RESULT 11

AAT68027

ID AAT68027 standard; DNA; 1731 BP.

XX AC AAT68027;

XX 15-JUL-1997 (first entry)

XX H. pylori flagella-associated protein ORF 07ge20415orf27.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacterium; life cycle; activator;

KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

KW diagnosis; ds.

XX Helicobacter pylori.

FH Key Location/Qualifiers

FT CDS 1..1731

FT /*tag= a

FT /note= "no stop codon given"

FT WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

XX P-PSDB: AAW20774.

XX Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

XX Claim 1; Page 854; 1481pp; English.

XX This sequence encodes a H. pylori flagella-associated protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX
 XX
 SQ Sequence 1731 BP; 600 A; 302 C; 384 G; 445 T; 0 other;

```
Query Match      82.2%; Score 14.8; DB 18; Length 1731;
Best Local Similarity 88.9%; Pred. No. 95;
Matches 16: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 acatcaagtgattcgcg 18
| | | | | | | | | |
pb 422 acatcaaacctcattcgcg 439

RESULT	12
AAF28548	
ID	AAF28548 standard; DNA; 96109 BP.
XX	
AC	AAF28548;
XX	
DT	04-APR-2001 (first entry)
XX	
DE	Genomic fragment #35.
XX	
KW	Genomic library; bacteria; human upper airway; otitis media; sinusitis;
XX	bronchopulmonary; endocarditis; meningitis; ss.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO200078968-A2.
XX	
PD	28-DEC-2000.
XX	
PF	16-JUN-2000; 2000WO-US16649.
XX	
PR	18-JUN-1999; 99US-0140121.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Lagace RE, Patterson C, Berg KL;
XX	
DR	WPI; 2001-041427/05.
XX	
PT	Genomic library for identifying diagnostic and therapeutic
PT	compositions, and for identifying virulence factors, regulatory
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic
PT	acids -
XX	
PS	Claim 1; Page 345-368; 545pp; English.
XX	
CC	The present invention relates to a Moraxella catarrhalis genomic library
CC	comprising of a combination of 41 nucleic acid molecules (see
CC	AAF28514-AAF28554). The library has a number of uses described in the
CC	specification e.g. is useful for identifying diagnostic and therapeutic
CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC	aerobic, gram-negative diplococcus, normally found among the bacterial
CC	flora of human upper airways. M. catarrhalis is known to cause acute,
CC	localised infections such as otitis media, sinusitis and bronchopulmonary
CC	infection and life-threatening, systemic diseases including endocarditis
CC	and meningitis.
XX	
SQ	Sequence 96109 BP; 287883 A; 18910 C; 20341 G; 28075 T; 0 other;

```

Query Match      82.28;   Score 14.8;   DB 22;   Length 96109;
Best Local Similarity 88.9%;   Pred. No. 1.7e+02;
Matches 16;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0

```

RESULT	13
AAF96936	
ID	AAF96936 standard; DNA; 21 BP.
XX	
XX	AAF96936;
XX	
XX	06-JUN-2001 (first entry)
DT	
XX	
DE	Human gene single nucleotide polymorphism #1697.
XX	
XX	Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
KW	polymorphism; vascular disease; coronary artery disease; forensics;
KW	myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
KW	pulmonary embolism; paternity test; ds.
XX	
OS	Homo sapiens.
XX	
XX	
FH	Key
FT	Variation
FT	Location/Qualifiers
FT	replace(11,G)
FT	/*tag- a
FT	/standard_name= "single nucleotide polymorphism"
XX	
PN	W0200118250-A2.
XX	
XX	15-MAR-2001.
PD	
XX	
PF	07-SEP-2000; 2000WO-US24503..
XX	
PR	10-SEP-1999; 990US-015457.
PR	26-JUL-2000; 2000US-0220947.
PR	16-AUG-2000; 2000US-0225724.
XX	
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Lander ES, Garjil M, Ireland JS, Holk S, Daley GO, McCarthy JJ;
XX	
DR	WPI; 2001-226749/..
XX	
PT	Nucleic acids comprising single nucleotide polymorphisms, useful in
PT	applications such as forensics, paternity testing, medicine, genetic
PT	analysis and phenotype correlations to diseases such as diabetes and
PT	atherosclerosis -
XX	
PS	Examples; Page 161; 242pp; English.
XX	
CC	The present invention provides a method of diagnosing a vascular disease
CC	in an individual, involving determining the sequence at various
CC	polymorphic sites within the human thrombospondin 1 and thrombospondin
CC	genes. The sequences at a number of polymorphic sites are also provided
CC	in the specification. In particular, the method can be used in the
CC	diagnosis of atherosclerosis, myocardial infarction, coronary heart
CC	disease, stroke, peripheral vascular diseases, venous thromboembolism
CC	and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also
CC	useful in forensics, paternity testing, genetic analysis and phenotype
CC	correlations to diseases. The present sequence is an example of one of
CC	the human gene SNPs shown in the specification.
XX	
SQ	Sequence 21 BP; 7 A; 3 C; 6 G; 5 T; 0 other;

Query Match 80.0%; Score 14.4; DB 22; Length 21;

Sequence 96109 BP: 28783 A: 18910 C: 20341 G: 28075 T: 0 other:

Best Local Similarity 93.8%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17
||| ||||| ||||| ||
Db 1 catcaaaagtattgcgc 16

RESULT 14
AAA48232/c
ID AAA48232 standard; cDNA; 69 BP.

XX AC AAA48232;

XX DT 28-SEP-2000 (first entry)

XX DE T. reesei xylanase, xYtv-104, oligonucleotide.

XX KW Xylanase; animal feed; digestion efficiency; thermostable;

XX KW feed pelleting; enzyme; xYtv-101; xYtv-102; Trx-103; xYtv-104;

XX KW xYtv-105; xYtv-106; xYtv-107; Trx-108; xYtv-109; xYtv-110; Trx-1;

XX KW xYtv-2; Trx-3; Trx-6; Trx-8; xYtv-4; xYtv-5; xYtv-7; xYn A; xYn B;

XX KW xYn; xYn C; xYn I; xYn II; ss.

XX OS Trichoderma reesei.

XX PN WO200029587-A1.

XX PD 25-MAY-2000.

XX PF 16-NOV-1999; 99WO-CA01093.

XX PR 16-NOV-1998; 98US-0108504.

XX PA (IOGE-) IOGEN CORP.

XX PI Sung WL, Tolan JS;

XX DR WPI; 2000-387799/33.

XX PT Thermostable xylanases useful for preparing animal feeds especially
XX PT poultry or swine feed, exhibits optimal activity under physiological
XX PT conditions -

XX PS Example 1; Page 74; 86pp; English.

XX CC Xylanase enzymes are added to animal feeds to increase the efficiency of
XX CC digestion and assimilation of nutrients. Xylanases are preferentially
XX CC added during the feed pelleting process. To survive the pelleting
XX CC process and to have optimum activity in the animal, the xylanase needs to
XX CC have high thermostability, with optimum activity at physiological pH and
XX CC temperature. Various xylanases have some properties suitable for feed
XX CC applications but they lack the thermostability required to survive
XX CC food pelleting. The present sequence is oligonucleotide, xYtv-104, from
XX CC Trichoderma reesei. This sequence was used along with 17 other
XX CC overlapping oligonucleotides to construct the coding sequence for

XX CC Trichoderma reesei xylanase Trx. The other oligonucleotides were:

XX CC xYtv-101 (AAA48229), xYtv-102 (AAA48230), Trx-103 (AAA48231), xYtv-105

XX CC (AAA48233), xYtv-106 (AAA48234), xYtv-107 (AAA48235), Trx-108 (AAA48236),

XX CC xYtv-109 (AAA48223), xYtv-110 (AAA48237), Trx-1 (AAA48238), xYtv-2

XX CC (AAA48239), Trx-3 (AAA48240), xYtv-4 (AAA48241), xYtv-5 (AAA48242), Trx-6

XX CC (AAA48243), xYtv-7 (AAA48244) and Trx-8 (AAA48245). The resulting Trx

XX CC sequence was mutated to produce mutant xylanases with enhanced

XX CC thermostability (AAV99683, AAV99684, AAV99685, AAV99686, AAV99735 and

XX CC AAV99736) which would be useful for animal feeds, especially poultry and

XX CC swine feed.

XX SQ Sequence 69 BP; 19 A; 15 C; 17 G; 18 T; 0 other;

Query Match 80.0%; Score 14.4; DB 21; Length 69;

Best Local Similarity 93.8%; Pred. No. 1e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17
||| ||||| ||||| ||
Db 34 CATTAAGTATTGCGC 19

RESULT 5

AAV36098/c
ID AAV36098 standard; DNA; 573 BP.

XX AC AAV36098;

XX DT 02-SEP-1998 (first entry)

XX DE DNA sequence of the specification.

XX KW Family 11 xylanase; improve; thermophilicity; alkalophilicity;

XX KW thermostolerance; bleach; wood pulp; processing; wheat; maize;

XX KW digestibility-improving animal feed additive; starch production;

XX KW mutant; ds.

XX OS Synthetic.

XX PN EP828002-A2.

XX PD 11-MAR-1998.

XX PF 05-SEP-1997; 97EP-0115412.

XX PR 09-SEP-1996; 96US-0709912.

XX PA (CANA) NAT RES COUNCIL CANADA.

XX PA (NARE-) NAT RES COUNCIL.

XX PI Ishikawa K, Sung WL, Yaguchi M;

XX DR WPI; 1998-161100/15.

XX PT Modified xylanase enzymes - useful for improving wood pulp

XX PT bleaching, etc.

XX PS Disclosure; Page 50; 84pp; English.

XX CC The present sequence appears in the specification, which describes a
XX CC method for modifying a Family 11 xylanase to improve its thermophilicity,
XX CC alkalophilicity and/or thermostolerance. This method comprises

XX CC modification of amino acids 16, 27 or 29 of Trichoderma reesei xylanase,

XX CC it or corresponding oligonucleotide sequences in the N-terminal region

XX CC replacement of one or more amino acid sequences in the N-terminal region

XX CC with corresponding aligned sequences from another Family 11 xylanase to

XX CC form a chimeric xylanase and/or upstream extension of the N terminus by

XX CC addition of up to 10 amino acids. The modified xylanases are useful for

XX CC improving the bleachability of wood pulp by treatment at 55-75 degrees

XX CC Celsius and pH 7.5-9.0 for 5-180 minutes. They might also be useful as

XX CC digestibility-improving animal feed additives. They might also be useful

XX CC in the processing of wheat or maize for starch production.

XX SQ Sequence 573 BP; 163 A; 127 C; 136 G; 147 T; 0 other;

Query Match

Best Local Similarity 80.0%; Score 14.4; DB 19; Length 573;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 catcaaaagtattgcgc 17
||| ||||| ||||| ||

Db 471 CATTAAGTATTGCGC 456

Search completed: April 2, 2002, 22:49:16

Job time: 9216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:46:04 ; Search time 2977.34 Seconds
(without alignments)
99.736 Million cell updates/sec

Title: US-09-785-904-4

Perfect score: 18
Sequence: 1 catgccgtcgtattgctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
C 1	18	100.0	650	6	AR096715	AR096715 Sequence
C 2	18	100.0	650	6	AR147696	AR147696 Sequence
C 3	18	100.0	1535	6	AR096713	AR096713 Sequence
C 4	18	100.0	1535	6	AR147694	AR147694 Sequence
C 5	18	100.0	1604	1	MBU35021	U35021 Mycobacteri
C 6	18	100.0	3453	1	MTU43540	U43540 Mycobacteri
C 7	18	100.0	9281	1	MBDR3S1	U35017 Mycobacteri
C 8	18	100.0	9764	1	AE006921	AE006921 Mycobacte
C 9	18	100.0	10019	1	AF041819	AF041819 Mycobacte
C 10	18	100.0	12412	6	186264	186264 Sequence 18
C 11	18	100.0	17783	1	AE007028	AE007028 Mycobacte
C 12	18	100.0	32437	1	MTCY336	295586 Mycobacteri
C 13	18	100.0	38380	1	MTCY251	274410 Mycobacteri
C 14	18	100.0	43401	1	MTV13E12	295390 Mycobacteri
C 15	16.4	91.1	9600	14	AF176573	AF176573 Hepatitis
C 16	16.4	91.1	10096	1	AF026197	AF026197 Xanthomon
C 17	16.4	91.1	19304	1	AE045311	AE045311 Xanthomon
C 18	16	88.9	11811	1	AE004334	AE004334 Vibrio ch
C 19	15.4	85.6	480	3	AF051144	AF051144 Mamestra
C 20	15.4	85.6	924	1	AF208062	AF208062 Neisseria
C 21	15.4	85.6	925	1	AF208061	AF208061 Neisseria
C 22	15.4	85.6	1779	6	A96206	A96206 Sequence 23
C 23	15.4	85.6	3712	1	NM065788	U65788 Neisseria m
C 24	15.4	85.6	4862	1	AF121135	AF121135 Neisseria
C 25	15.4	85.6	4883	1	AF313394	AF313394 Neisseria
C 26	15.4	85.6	5859	1	NGU14554	U14554 Neisseria g
C 27	15.4	85.6	5859	6	AR025350	AR025350 Sequence
C 28	15.4	85.6	5859	6	AR025351	AR025351 Sequence
C 29	15.4	85.6	5859	6	AR111460	AR111460 Sequence
C 30	15.4	85.6	5859	6	124768	124768 Sequence 1
C 31	15.4	85.6	5859	6	124769	124769 Sequence 7
C 32	15.4	85.6	83921	9	AP001255	AP001255 Homo sapi
C 33	15.4	85.6	97811	2	AC019980	AC019980 Drosophil
C 34	15.4	85.6	126253	8	AC019018	AC019018 Arabidops
C 35	15.4	85.6	153865	2	AC091965	AC091965 Homo sapi
C 36	15.4	85.6	163687	3	AC008200	AC008200 Drosophil
C 37	15.4	85.6	174637	2	AC027490	AC027490 Homo sapi
C 38	15.4	85.6	177127	3	AC008199	AC008199 Drosophil
C 39	15.4	85.6	23313	3	AE003741	AE003741 Drosophil
C 40	15.4	85.6	340000	9	AP001686	AP001686 Homo sapi
C 41	15	83.3	439	14	AB030969	AB030969 Hepatitis
C 42	15	83.3	742	6	E06698	E06698 non A non B
C 43	15	83.3	742	6	149752	149752 Sequence 18
C 44	15	83.3	742	6	163390	163390 Sequence 18
C 45	15	83.3	932	14	S46012	S46012 polypeptide

ALIGNMENTS

RESULT 1	AR096715/c	Sequence 5	650 bp	DNA	PAT	08-SEP-2000
LOCUS	AR096715	Sequence 5	from patent US 6008201.			
DEFINITION	AR096715					
ACCESSION	AR096715.1	GI:10025753				
VERSION						
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 650)					
AUTHORS	Riley, L.W.					
TITLE	DNA molecule encoding for cellular uptake of mycobacterium tuberculosis and uses thereof					
JOURNAL	Patent: US 6008201-A 5 28-DEC-1999;					
FEATURES	Location/Qualifiers					
source	1..650					
BASE COUNT	123 a	258 c	185 g	84 t		
ORIGIN						

```
Query Match      100.0%; Score 18; DB 6; Length 650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
|||||
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 2
ARI147696/c
LOCUS      ARI147696      650 bp      DNA      PAT      08-AUG-2001
DEFINITION Sequence 5 from patent US 6224881.
ACCESSION  ARI147696
VERSION     ARI147696.1 GI:15111786
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 650)
AUTHORS    Riley,L.W. and Chong,P.
TITLE      DNA molecule fragments encoding for cellular uptake of
            Mycobacterium tuberculosis and uses thereof
JOURNAL    Patent: US 6224881-A 5 01-MAY-2001;
FEATURES   Location/Qualifiers
            source
            1..650
            /organism="unknown"
BASE COUNT 123 a 258 c 185 g 84 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 650;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
|||||
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3
AR096713/c
LOCUS      AR096713      1535 bp      DNA      PAT      08-SEP-2000
DEFINITION Sequence 1 from patent US 6008201.
ACCESSION  AR096713
VERSION     AR096713.1 GI:10025749
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1535)
AUTHORS    Riley,L.W.
TITLE      DNA molecule encoding for cellular uptake of mycobacterium
            tuberculosis and uses thereof
JOURNAL    Patent: US 6008201-A 1 28-DEC-1999;
FEATURES   Location/Qualifiers
            source
            1..1535
            /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 1535;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057
```

```
RESULT 4
ARI147694/c
LOCUS      ARI147694      1535 bp      DNA      PAT      08-AUG-2001
DEFINITION Sequence 1 from patent US 6224881.
ACCESSION  ARI147694
VERSION     ARI147694.1 GI:15111784
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1535)
AUTHORS    Riley,L.W. and Chong,P.
TITLE      DNA molecule fragments encoding for cellular uptake of
            Mycobacterium tuberculosis and uses thereof
JOURNAL    Patent: US 6224881-A 1 01-MAY-2001;
FEATURES   Location/Qualifiers
            source
            1..1535
            /organism="unknown"
BASE COUNT 297 a 544 c 458 g 236 t
ORIGIN

Query Match      100.0%; Score 18; DB 6; Length 1535;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 5
MBU35021/c
LOCUS      MBU35021      1604 bp      DNA      BCT      27-APR-1996
DEFINITION Mycobacterium bovis BCG DNA flanking deletion region 3.
ACCESSION  U35021
VERSION     U35021.1 GI:1049243
KEYWORDS   .
SOURCE     Mycobacterium bovis BCG.
ORGANISM   Mycobacterium bovis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE  1 (bases 1 to 1604)
AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE      Molecular analysis of genetic differences between Mycobacterium
            bovis BCG and virulent M. bovis
JOURNAL    J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE    96200095
REFERENCE  2 (bases 1 to 1604)
AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
            Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES   Location/Qualifiers
            source
            1..1604
            /organism="Mycobacterium bovis"
            /sub_species="BCG"
            /db_xref="taxon:1765"
            misc_feature 1400
            /note="site of deletion breakpoint in this organism.
            Virulent M. bovis has additional 10 kbp at this site."
BASE COUNT 337 a 637 c 416 g 214 t
ORIGIN

Query Match      100.0%; Score 18; DB 1; Length 1604;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
|||||
Db 698 CATGCCGTCGTATTGCTG 681
```

	/note="previously known as rfbC"
	/codon_start=1
	/transl_table=1
	/protein_id="AAB66649.1"
	/db_xref="GI:2326951"
	/translation="MKREPRHLDPGAGWEITPTIHVDSRGLFEMLTQHGCFKAFACGS LIVPQVNCSSVSCGVRLCHLP AOLPSQAKYVTCSGVDFDVVDIREGSPHGRWDS VLDDQDNKIYVSNGLAHGFALQDNTVMYLCSAEYNHQROHTICATUPTTASIRKMS SDWGCPSLSDRDAAAPSFEEIVRRGLIPGGTDVSRLFGEWRGT"
	2202...2655
	/note="3-4 copies in M. tuberculosis genome; Found in the rib cluster (inside the rfbA gene), also found in the mec of M. tuberculosis; present at 3-4 sites of the genome of H37Rv, H37Ra, Erdman, and many isolates; not found in M. Smegmatis & M. avium."
	/rpl_type-dispersed
	/rpl_unit=-1806...2250
	BASE COUNT 679 a 1226 c 990 g 558 t
	ORIGIN
	Query Match 100.0%; Score 18; DB 1; Length 3453;
	Best local Similarity 100.0%; Pred. No. 35;
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	Qy 1 catgccgcgtatctg 18
	Db 2448 CATGCCGCCTATTCTG 2431
	RESULT 7
	MDDR3S1/c
	LOCUS
	DEFINITION
	ACCSSION
	VERSION
	KEYWORDS
	SEGMENT
	SOURCE
	ORGANISM
	REFERENCE
	AUTHORS
	TITLE
	JOURNAL
	FEATURES
	source
	misc_difference replace(1400..>9281,"")
	/note="absent in related avirulent strain M. bovis HCC"
	BASE COUNT 1706 a 3191 c 2929 g 1453 t
	ORIGIN
	Query Match 100.0%; Score 18; DB 1; Length 9281;
	Best local Similarity 100.0%; Pred. No. 29;
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	Qy 1 catgccgcgtatctg 18
	Db 698 CATGCCGCCTATTCTG 681

RESULT 8
AE006921 9764 bp DNA BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 7 of 280 of the
DEFINITION complete genome.
ACCESSION AE006921 AE00516
VERSION AE006921.1 GI:13879142
SOURCE
ORGANISM Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 9764)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 9764)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
TITLE Submitted (25-APR-2001) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .9764
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
122. .1414
/gene="MT0095"
/note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; similar to SP:P16431 GB:X17506
PID:41684 PID:882614 GB:000096; identified by sequence
similarity; putative"
1449. .2123
/gene="MT0096"
1449. .2123
/gene="MT0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44319.1"
/db_xref="GI:13879143"
/translation="MSVYKHPSPRVLRQTRSTVVKGRSGSLSWRRVKRTGDLGLAVMG
GREERAVKPTGTPQPKDMMTVFVDAGPGRSVRSVEVAAPAELEAIADPADPRHR
ELDGSCTRGNIKVAKLVGSKFSTKMKLFGLPRTISRVTALPKNHLVCSHPHG
RWRWEPESLPTLTRVTETFDYHAAGATKNGLKLYEMTGFAKSNAAGITATLAKLSDQ
YARGRA"
2280. .2873
/gene="MT0098"
2280. .2873
/gene="MT0098"
/note="similar to GP:145427; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAK44320.1"
/db_xref="GI:13879144"
/translation="MDQPNANIHYDALLDMVPLTQCQVLVDGCGDGLIAAKLARRI
PYVTAYDIDAPVLRRAQTRFANPIRWLHADIMTAELPNAGDAVVVSNALCHHIEDTR
TALSRLGLVTPGGTLAVVTFTVPSLRNGLWLHTSWVACGMANKVRKGRWEHSPIKWP

gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene

PROTLHELRSIVRALLPAGCIRRLLYGRVLVTRAPV"
/gene="MT0099"
3002. .3772
/gene="MT0099"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNQNRINRWELITCGLGHHVYAPDDAALAANKRAS"TGLE
VMKRLRGDFALGPGOGAPEDAPLIMRGKALRQAIIRALGVERIVRALVLAAM
AVWFLRGARGAIOATIDKDIPLVRAAGFKVDQMTVTHALEKALAAKPSLALITGMILA
VYILQAVGEGWLMILKMGKRYFAVVATSIPLPLEVHOLAKGIITTRVVTFSINAAV
3791. .4162
/gene="MT0099.1"
3791. .4162
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MRGIRVVSVKVTSAFFAIFTPARRAKAARVNLQVAVOCCRKPDEG
PTLCOHCUPGSAARALPTAAWSKQSRVPATHCRPCCAPGAASAALTCALCAFAWSVVE
VR/APRAATESLQSKAAARESH"
4206. .4973
/gene="MT0100"
4206. .4973
/gene="MT0100"
/note="similar to GB:AL00912; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine
nucleosidase/S-adenosylhomocysteine nucleosidase,
putative"
/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTGVGICAIPOELAYLRGVLDVAKRQVAVQIILFDSQGLDAHR
VILAAAGMKVNTGILTALLADREGCTIVFTGVAGGLDPELCIGDIIVADRVQHDIF
GLHDERIKPTQPGHTIPHEITRELGYFVDVAVIDRVKRLDGFILAIPLSTAAGGGR
QRIYVYGFILTGQYLVIRTERPNRLHHELGMAVEMPGGAVQIICASDIIPLVITRAL
5105. .7490
/gene="MT0101"
5105. .7490
/gene="MT0101"
/note="similar to SP:P45840 PID:559912; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTTAVTGHHSVORIOLRISGMSCSACAHVESTLNKLPQVRA
AVNFGTRVATIDTSEAVIAALCQAVRRAGYQADLCTDDGSRASIPDADHARQILIKL
AIAAVLFVPVADLSYMFVGPVATPRTGQWVLSALALPVTAAAPFHRVARNARHH
AASMETLISVGITAAITWSLYTFGNHSPITERSGIQWALLGSDAIYFEVAGVTVFVL
VGRYFEAKAKSQAGSALAKAALSAKEVALLPDGSEMIIPADEIKEQRFVVRPQQI
VAAGLAVDGSAAVDMDSAMTGEAKPTRVPGQVIGGTVTLDGRIIVEAAAAGADPTOF
AGWRLVHQQAQKADARLADRISSVFPVAVIALTAAAGWLIAGGUPHVASAAL
AVLITACPLGLATPTAMVWASGRGAOLGIFLKGYKSLEATRAVTVVDPKTCGLTT
GRIQVSATVAAPGNEADQALAAATVAAASERHSVALATAATTRDAVTDKPIAPGRG
VSGTSGRAVRVPGKPSWIGSSCHPNRAARHAEISLGETAVFVDEGPGCVIADAV
AVKDSADAVAAALADRGRTMLLTGDNPESSAAVATRVGIDEVIADIILPEGRVDVIEQ
LHDGRHVAVMGDGINDCGPALARADLGNAGTGTVAICAADIIIVRHILDVVPLAID
LARATPVRVKNWNAFCYNTAAIPVAAAGLNLPLVAGAAAFSFEFVSVNSLKRKF
GHPUGCGTVGGPQMTAIPSSA"
complement(7337. .8185)

CDS	/gene="MT0102" complement(7337..8185) /gene="MT0102" /note="identified by Glimmer2: putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK44325.1" /db_xref="GI:13879149" /translation="MIAQAFTAGSPNIIHASTVLOGRCVPAAMWSEHPAGATIRHCATII DGMDCVEAREALSGEKAPVISAQVDEHLGECACKAMFTQVASQAGDLRRRLAES RPVVPVGRIGIRRAPRRQHSPTMRWALLCVGIAQIALGTVOGRGIDVGTHTGHT GACTHLLNTESTSIALGVIMVGAALWPSAAAGLAGVITAFVALITCYVIVIALSCAV STRIILLPLVVGAVLAIMVSSASQPRPRPDAAAEFDIIVIPDNASKRRRRGHLWP TDGSA"		Yu.S. Direct Submission Submitted (07-JAN-1998) Microbiology & Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10010, USA Location/Qualifiers 1..10019 /organism="Mycobacterium bovis BCG" /strain="Pasteur" /sub_species="bcg" /db_xref="taxon:34892" complement(259..1018) /note="ORF1: similar to MYCY336.37" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK96957.1" /db_xref="GI:2791843" /translation="MKRSKSGSFAGLAMLIAPVAAHGLAAADIAHPVYQQLTD VITARGISQGVTFSWAGGTSPTGRTGTINTVGFDSAGLITVAYAGAGLKLPHSS GMYKVKQVLPQAKKGDIFYPGPGTOSVALYLGKQMLEVLDVVVSPVETNMT PVLVGFSGPSRRKSRSSQRSSKRPSSKRGARITGAGITGAGIPASAPASAR PAASLHRAITLITVPLAFSTHPKPSACFPLLSARSRLVWVN" complement(1278..1850) /note="ORF2: similar to MYCY336.36" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK96955.1" /db_xref="GI:2791844" /translation="MICSQVNPVAAAMEIEGNAATPRIPVTVAKTSAGIETACMAP INQPRDIQDAGTESKSLIDEPDPRPRLTLRAPDVCIVVILAVELVAVPTTS WSPKPAFTDNCVPHACVLPVGVNYYIYPMWGGIGAAIAVILGVVSHLKGWYM SPWPISTAVITVTSTAGYATAGSESEKYWH"	
gene	/gene="MT0103" complement(8232..9095) /gene="MT0103" /note="identified by Glimmer2: putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK44326.1" /db_xref="GI:13879150" /translation="MDWLHPIDGITPTERARKGDTISNQYIDMSKLSGYLPQARA TFEAVLAKLAAPGATNPDIHTPVDTPTAAATIDROTQSAQRNHIITLILAGLRALIAS KGLQHGRLPVSIVVITTLTDLDTGAGKGTTCGTLTPMADVIRNTSHAHYSPASCR YQNFIDHGTPLALYHTKKLASPAQRIMLFANDRGCTKPKLDAPAVISQAHIVTANTS TGRDITTELTACGPONKLAERGWTHKNTGHGTWELPPHLLDQGRNTTFHHPEKF LHNQDDDDKPP" complement(8956..9623) /gene="MT0104" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact: identified by Glimmer2; putative; conserved hypothetical protein, authentic frameshift" BASE COUNT 1553 c 3360 g 1748 t		gene 1810..3124 /gene="bioA" /note="bioA: similar to MYCY336.35c" /codon_start=1 /transl_table=11 /product="putative adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AAK96956.1" /db_xref="GI:2791845" /translation="MAANTGULTPPGJLIVAGAILMHPYSSIGREAVSPVVAARAR WTLIRKQPTFELVLSASSWATAIIRHGHPIALQALITQLRVNHHVETGCTHEPAAR LAPLVLITPAGLIDVILFSGSSSVSAAKMALIYMGKGLERKKGLPTWRGKYCHD LITAMPCTDPRGTHGCVLVLAAQVFAIPVIRVDIYASATLADLADHAEALAAV VLIYVQAAQARHTHPTVLEKRLDITGRYEVLLIDFLATGCTLGAALFAADHATSP LIRVGAALGGTLELAAALITFALVANTISAGAGALMURPILRANILATAVASVAF LITQDARKITELAAALIGLIDITAKALPAVTVKVGAGATCVLITDHPVILAVATPAA LIDGVMIRPPEINLYAMPTTCTPAELTQTTSAMVEVARLVGSLP"	
gene	/gene="MT0104" complement(8956..9623) /gene="MT0104" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact: identified by Glimmer2; putative; conserved hypothetical protein, authentic frameshift" BASE COUNT 1553 c 3360 g 1748 t		gene 1810..3124 /gene="bioA" /note="bioA: similar to MYCY336.35c" /codon_start=1 /transl_table=11 /product="putative adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AAK96956.1" /db_xref="GI:2791845" /translation="MAANTGULTPPGJLIVAGAILMHPYSSIGREAVSPVVAARAR WTLIRKQPTFELVLSASSWATAIIRHGHPIALQALITQLRVNHHVETGCTHEPAAR LAPLVLITPAGLIDVILFSGSSSVSAAKMALIYMGKGLERKKGLPTWRGKYCHD LITAMPCTDPRGTHGCVLVLAAQVFAIPVIRVDIYASATLADLADHAEALAAV VLIYVQAAQARHTHPTVLEKRLDITGRYEVLLIDFLATGCTLGAALFAADHATSP LIRVGAALGGTLELAAALITFALVANTISAGAGALMURPILRANILATAVASVAF LITQDARKITELAAALIGLIDITAKALPAVTVKVGAGATCVLITDHPVILAVATPAA LIDGVMIRPPEINLYAMPTTCTPAELTQTTSAMVEVARLVGSLP"	
Query Match	100.0%; Score 18; DB 1; Length 9764;		AF041819 10019 bp DNA BCF 18-JAN-1998	
Best local	100.0%; Pred. No. 29;		Mycobacterium bovis BCG putative	
Matches	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		adenosylmethionine-8-amino-7-oxononanoate aminotransferase (bioA), putative 8-amino-7-oxononanoate synthase (bioF), putative dethiobiotin synthetase (biob), and biotin synthetase (biob) genes, complete cds.	
QY	1 caatgcgcgaattgctg 18 		AF041819 AF041819.1 GI:2791842	
Db	9003 CATGCCGCGATTCTGCTG 9020		Mycobacterium bovis BCG. Mycobacterium bovis BCG Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)	
RESULT	9		AF041819	
LOCUS	AF041819 10019 bp DNA BCF 18-JAN-1998		Mycobacterium bovis BCG putative	
DEFINITION	adenosylmethionine-8-amino-7-oxononanoate aminotransferase (bioA), putative 8-amino-7-oxononanoate synthase (bioF), putative dethiobiotin synthetase (biob), and biotin synthetase (biob) genes, complete cds.		AF041819 AF041819.1 GI:2791842	
ACCESSION	AF041819		Mycobacterium bovis BCG.	
VERSION	AF041819.1		Mycobacterium bovis BCG	
KEYWORDS	Mycobacterium bovis BCG. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)		Mycobacterium bovis BCG. Mycobacterium bovis BCG Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)	
ORGANISM	Mycobacterium bovis BCG.		Mycobacterium bovis BCG	
REFERENCE	1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)		Mycobacterium bovis BCG. Mycobacterium bovis BCG Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)	
AUTHORS	Yu.S. and Jacobs,W.R. Jr.		Mycobacterium bovis BCG.	
TITLE	Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)		Mycobacterium bovis BCG Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 10019) Yu.S. and Jacobs,W.R. Jr. Cloning, sequencing, and identification of Mycobacterium bovis BCG biotin biosynthetic genes by complementing two Mycobacterium smegmatis biotin mutants Unpublished 2 (bases 1 to 10019)	
JOURNAL	Unpublished		Mycobacterium bovis BCG	
REFERENCE	2 (bases 1 to 10019)		Mycobacterium bovis BCG	

AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Whole genome comparison of mycobacterium tuberculosis clinical and laboratory strains
JOURNAL REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 17783) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE	Direct Submission
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers 1..17783 /organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" complement(104..826) /gene="MT1622.1" complement(104..826) /gene="MT1622.1" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45890.1" /db_xref="GI:13881251" /translation="MLAKLAAGATNFDDHTPTVIDTTPDAAA1DRDTRSOAQRNHGCL LAGRLIASKLGQHNGLPVSIVVTTLTDLQAGAGKFTGGSTLLPMADVIRMTSH AHHSPASGRTPQIFDHFGLPTLALYHKRLASQAQRIMLEANDRGCTKPGCCDAPAYIS QAHVHTAWTSGRTDITDLTLACDDPNRLAEKGTWTKHNTGHITELWLPPLHLDHGQPR TNPFHHEKLRHNDENDDHP" complement(831..1499) /gene="MT1623" complement(831..1499) /gene="MT1623" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45891.1" /db_xref="GI:13881252" /translation="MLANSREELVEVFDALDAELDKLDEVSEFVLTTPFKLSLEKLE CLVRLPVGHTLLNQDTONASEEELGGTLCCLAIANKLRTKTPAALRIADAAIDGHR RALTCPLAPQLTATATAOQGLTGEAHRVIRALFPPARRKGGVHHPPGRSRKPGHQ SKTSISSRRAGPLRPAGHGLATPRKRPHIRTRPQTRHIEQAIRKHVTAKWLPDPIIS AGHL" 1947..2996 /gene="MT1624" 1947..2996 /gene="MT1624" /note="similar to GP:1680650; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="biotin synthase" /protein_id="AAK45892.1" /db_xref="GI:13881253" /translation="MTQAATRTPDAGQDGNNSDILVVARQOVLQRGEGINQDVILA VLOLPDDRLELLALAEVRMRCGPEVEVGGIISLKTGCPDCHFCSSGSLFASVP RSNWIDPLSVEAKQYAKSCATFECIVAAVRGPDRLMAQVAAGIEAINEVEINIA CSGLGMLTREQVDQLAAGVHYRNINLETARSFANVTTHTEERWOTLSMWADGME VCCGGLTQAGETLOORAEEFAELAEGLPDEVLNPLNPRCTPPADLEVPVPGDALKAA VAFRLALPRLMRLFAGGRETITGLDGLAKRGILGGINAVIVGNLTJTLGRPAEADLEL LDELQMLPLKALNASL" 2997..3236 /gene="MT1625"
gene	CDS 2997..3236 /gene="MT1625" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45893.1" /db_xref="GI:13881254" /translation="MVEIVAGKORAIHVAAGVYNYVTGELADTATPTAARMGLEPPRFCAACGRMVMYQVRPDDGNWAKSRHQVDSADLATQR" 3117..3898 /gene="MT1626" 3117..3898 /gene="MT1626" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45894.1" /db_xref="GI:13881255" /translation="MIGISATGVLVGLWMIAPPPIHAVVAITRAGERVHYHIGSESQ NFFIAPPMLIGLISVLAVVASALMOWREHGRGPOMVAGISIGITTTAAAIAGVCAIVV RLYGALDFDTVPISKGDIALTYVTQAPVFFARRPLQIALTIMPAGIASIVYALLA AWTARDLGGYPAYDPPSSNARTEALETFQAPVS" complement(3953..4054) /gene="MT1627" complement(3953..4054) /gene="MT1627" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45895.1" /db_xref="GI:13881256" /translation="MKSATHGANSDLAHLNLTVDLNPQDDPLESKR" complement(4063..5430) /gene="MT1628" complement(4063..5430) /gene="MT1628" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45896.1" /db_xref="GI:13881257" /translation="MPDHGGISLMVPGNLAGATGAEMGRDPHEELQRKVRPLPBSD DPKYFPAGYHAPVTVLRSKDELAEMGLIPQPVATATQLLYRTTMVGNPEATVTT VIVPAMALAPGCTLLSYGCAIDAMSSRCFPYALRRRAKALSSITQMELIMISAALA EIVAVSPVPHGGPKHLWGSHYEPGYHVLGDLRAALNSRVKGLSPATPGLWGYSGGGL AASAAAFACVCAVATDIDVAVIGSHVGLGHTFERKINCTLLALPALVVAALQVSYLP GLAKVIRKAMHGRGRLTGLDPMPTIVDAVIRMAKUMGDFLHPELLELSTPEVSHV LPHDKLASAPVPPVYLLVQAVHDTLIDVSDIDALADSYTAGGANVTYHRDLTSEHVSLL HPLSAPMTLRLWLTDRFAKRLTDRHVRVTWPTTFENPMYAGMARLAVIAAKVITGRKKL SRRKPL" complement(5660..6193) /gene="MT1629" complement(5660..6193) /gene="MT1629" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK45897.1" /db_xref="GI:13881258" /translation="MTSSVVRKLAFAKVDRELAHLFQAVFSDPHRLPGIRMIASITYL GVVPSPATPELPADTRPSSKSLPPNMFADHGPMTVHARTRLAKMSYTNIGALAPKE FALSTLRDIYGAALGYQVDTATNLQRLVLRARRVITQGTITAQSGRSGRPAALYRPTDS QLRTVDEEALRPQGQL" 6419..7468 /gene="MT1630" 6419..7468 /gene="MT1630" /note="similar to SP:P11458 GB:X12713 PID:581137 GB:000096

PID:1651334; identified by sequence similarity; putative"

/codon_start=1
/transl_table=11
/product="quinolinate synthetase"
/protein_id="AAK45898.1"
/db_xref="GI:13881259"
/translation="MTVNLNRDTLDVDELADITNTPLGCVGVDCGRWAAEIRRLAHL
RGATVLAHNTQLPAIQADVHDGSDLSALSRVAAEAPEDTIVFCGVHFMATETAKILSPH
KTVLIPDQAGCSLADISTDELRAWKDEHPGVVSYNTTAAVKALTDICTSSNA
VDVVASIDPQREVLPCDFQFLGAVRRTGKLNHYWAGBCHVHAGINGDELADQARA
HPDELVHPHCEGCATXALAGEGAPFAERVKILSTGGMLEAAHTTRAROVLAPEV
GMLHOLRAAPEVDFAVRANDRASCKYMKITPAALLRCLVEGADEVIVDPGTAASCR
SVORMIEGHPGGE"
7468..9051
/gene="MT1631"
7468..9051
/gene="MT1631"
/note="similar to GP:608530; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AAK45899.1"
/db_xref="GI:13881260"
/translation="MAGPARDADVVVIGTVAGLAAALAADPAGSVVVLKAAQT
HVTATHYAQGGIAVVLDPNDSDVDHADVADTLAAGAGLCPDPAVYSIWAICYRAVTDLV
CAGARLDESVPGRWALTRREGHSRRIRVHAGDGATGAEVORLQDAAGMLDITGTHVA
LRVLDGTAVTGLLVVRPDGCIISAPSVLATGGLGLHYSATTNPAGSTGDTALGSL
WAGVASDLETFIQFPTMLFAGRAGSRPLITEAIREGAILVDROGNSITAGVHPMG
DLAPRDVVAIDARLAKTGDPCVYLDAGEVARTGLHGANRNASNLLEGIVGGGRAG
VYPAHYSVCGGIVTDVYGOTELGLVAAGEVARTGLHGANRNASNLLEGIVGGGRAG
KAAAHAAAGRSRATSSATWPEPISVYALDRGDLORAMSRDASMYKAAAGLHRLCDS
LSGADVQVACRRQFEDVALTILVQSVTAALARTESRGCHHRAEYPCTVPEQARSIV
VRGADANAVCQVALVAVC"
9021..9908
/gene="MT1632"
9021..9908
/gene="MT1632"
/note="similar to GB:L20833 SP:P30011 GB:L28105 PID:310262
PID:456041; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
Query Match 100.0%; Score 18; DB 1; Length 17783;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 catgcegcgtattgctg 18
|||||
Db 878 CATUCCGTCGTATTGCTG 895
RESULT 12
MTCY336
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 32437)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
Tekaita,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltwell,I.T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squires,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and

Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)
98295987
2 (bases 1 to 32437)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
on Jun 27, 1998 this sequence version replaced gi:211723.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
1..273
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
/clone="Y48"
33..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
/clone="Y336"
complement(39..2224)
/gene="Rv1565c"
complement(39..2224)
/gene="Rv1565c"
protein, some similarity to 005402 HYPOTHEICAL 72.2 KD
protein from H. subtilis (634 aa) opt: 384 E(); 4-8e-17;
29.1% identity in 378 aa overlap and to hypothetical
protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, Y392.HAEIN P43993 hi0392 (245 aa) opt: 265;
E(): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse
score is 0.930
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1565c"
/protein_id="CA809072.1"
/db_xref="GI:2117272"
/db_xref="SPTREMBL:O06625"
/translation="MLTISPPRPALTPEPALPPVTMGTRTTGFRHIDLKLRGVAIA
LVVPHVWFGVSGVDVFLALSGFFEGKTLRAALNDLSLPIAEVIRKLRLLPA
HTWMSVQGFYLAFLLVAGCAGLTLRLFRGPRAPYRTMFVFLVLLSTLTFLASHLYAI
VAHAYQATAYNTFAWELLAGALVGVVHPVPMWMLTAVATAALAIILSCGAL
IDGVKEFFGPWALVPVGTMLMAGANRGVPGTRDRPLNPLATAPILVILGAMA
YSWLHWPPLLIEMLSYTGHRHANFEAGVLLNSGLLAYLTRVEDPLRPAFAGV
RSPAAYPIPHRLRLRPTILVSGSVALLGVALTATSTFWREHIVLQRAAKELSGLS
SHDYPGARALIDHVRVKLMRPRTVLEVRHDLPTSKDGCISDFVNPALINCTYGDVD
APRTTALAGGSHAEHMLTALDLLGRMHFKVVTYLLKMGCPSTEEVPIIMGNAPYPO

```
/gene="bioA"
/note="PS00600 Aminotransferases class-III
pyridoxal-phosphate attachment site"
5101..6261
/gene="bioF"
5101..6261
/gene="bioF"
/note="Rv1569, (MTCY336.34c), bioF, len: 386. Function:
probable 8-amino-7-oxononanoate synthase (EC 2.3.1.47)
006621. Contains aminotransferases class-II
pyridoxal-phosphate attachment site (PS00599). FASTA
results: BioF_MYCIE P45487(385 aa) opt: 1971; E(): 0;
80.1% identity in 381 aa overlap. Also similar to
MTCY10H4.32, FASTA score: E(): 5.5e-29; 37.4% identity in
393 aa overlap"
/codon_start=1
/transl_table=11
/product="bioF"
/protein_id="CAH09068.1"
/db_xref="GI:2117268"
/db_xref="SWISS-PROT:O06621"
/translation="MKAATQARIDDSPLAMLDVAVQQRHEAGLRCLRRPRPAVATELD
LASNDYLGSRHPAVIDGVALRIWAGATGSRVLTGDKLHQFEEALAFVGAAD
GILFSGYFANI GAVVGISGIGSLIVSOARSHASIVDACLSKRVVVTTHKDVDAVD
AALRSDEKAVVVTDSVFSADGSLAPVRELLEVCVKHICALIIVDEAHGKGVKGGRG
LILFEGLAGAPVAVMTTFLSKALGSGGVVLGTPVRAHLIDAARPFIDFTVLAIAAV
GAARALRVLDQAEPRHPQAVINHAGELARMCGVAAPDASAMVSVIILGESESAAAA
CIIAGVKVCGFPPPTVAGTSRLRLTARASLNAGELELARRVLTDLAVARK"
/gene="bioF"
5806..5835
/note="PS00599 Aminotransferases class-II
pyridoxal-phosphate attachment site"
6258..6938
/gene="bioD"
6258..6938
misc_feature
Query Match 100.0%; Score 18; DB 1; Length 32437;
Best local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 catgcatgcatgattatcttq 18
|||||
Db 17614 CATGCCGTCGTATGCTG 17631
RESULT 13
MTCY251
LOCUS MTCY251 38380 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 5/162.
ACCESSION Z74410.1 AL121456
VERSION Z74410.1 GI:3261600
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis
Hactelia; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 38380)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.B.,
Tekaia, F., Badcock, K., Hain, D., Brown, D., Chillingworth, P.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellon, S.,
Squires, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature, 393 (6685), 537-544 (1998)
JOURNAL 98295987
MEDLINE
REFERENCE 2 (bases 1 to 38380)
AUTHORS Parkhill, J.
```

```
CHQVQAAAKLVADHPDVFVFTSTRPMNKKPGDVMPTATVVGWTQTFADNNIPVLAMR
DTPQLVKDGPETPADCLAKGGNQSCGIARSKVLDRNPRTLDVFARFPLKPLDMSD
AICRTDCRAVEGNVLVYRDSHLLTPYMRWTSETSELGROIAANTDW"
complement(2327..3019)
/gene="Rv1566c"
complement(2327..3019)
/gene="Rv1566c"
/note="Rv1566c, (MTCY336.37), len: 230. Function: unknown
probably exported has OAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasins
1.2(007390, 007391) from M. avium and slightly similar to
C-terminus of Listeria invasion-associated protein p60
precursor P60_LISMO P21171. FASTA results, Q49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. Tbpase score is 0.890"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAB09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:O06624"
/translation="MKRSKSGSFAIGLAMLAPMVAAGIAAADPATRPVDYQQITD
VVIARGLSORGVPSWAGGIGSGTRGTGTGINTVGFDSGLIQYVAGAGLKLPRSS
QOMTKVQKVLPOARKGDLLFYGPECTQSVALYLCKGOMLEVGDVVQVSPVPTNGMT
PYLVKRVGTQPTPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAP
VQPPFGTARSR"
complement(3030..3036)
/note="possible RBS, AAGAGG, for Rv1566c"
complement(3258..3263)
/note="ASNI site: ATTAT; probably linking fragments
B3/G6"
complement(3259..3543)
/gene="Rv1567c"
complement(3259..3543)
/gene="Rv1567c"
/note="Rv1567c, (MTCY336.36), len: 94. Unknown. Tbpase
score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:O06623"
/translation="MVTMTSPSLRFAFTDNCVCPDACPVPFCVNYIYIPVMWGGIG
AAIATAVIGPVSMLKGWYMSFWPIISIAVITVTSIAGYAIAGFSEKIWH"
3791..5104
/gene="bioA"
3791..5104
/gene="bioA"
/note="Rv1568, (MTCY336.35c), bioA, len: 447. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC2.6.1.62) 006622. Contains
aminotransferases class-III pyridoxal-phosphate attachment
site (PS00600). FASTA results, BIOA_MYCIE P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E(): 3.5e-16; 29.5%
identity in 421 aa overlap. Tbpase score is 0.874"
/codon_start=1
/transl_table=11
/product="bioA"
/protein_id="CAB09069.1"
/db_xref="GI:2117269"
/db_xref="SWISS-PROT:O06622"
/translation="MAATAGTLPETQIIAVDGAHLHPYSSIGREAVSPVVAHAHGA
WLTIRDGQPIEVLDAAMSSWMTAIHGHGHPALDQALTTQLKMNHVMFGGLTHEPAR
LAKLVITDPAGLDTVPFSDGSVSVEAAKMAQYWRGRLPGKRLRMWRRGYYHGD
TFLAMSITCDPHGHSHLWTDVLAQVAPQVPDYPDAYSAFPAQLAQHAGLAAVY
VEPVQAGRRFHDPRYLHDLRIDCRREYVLLIFDETATGFRGTGALFADHAGVSP
DIMCVGKALTGGYLSLAATLCTADVAHTISAAACALMHGPTENANPLACAVSVASVE
LLDGDWRTRITELSLAAGLTAGLDRALPALPVDVRCVCAIGVIECDRPDLAVATPAA
LDRGVWLRRPFRNLVTAMPPIYCTPAETITQITSAMVEARLVGSLP"
4541..4654
```

misc_feature


```

/codon_start=1
/product="hypothetical protein Rv3449"
/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/translation="MTTSRTLRLVLSALATSLGLGTPVAHAVSPPIIDRWLPESAL
PRLPPVQREVCETVEAESGRAPRAERSAQLADLDQVWRLTRGQORVAVIDTGVA
RHRPLKVGAGDVFTGDTGADCAHGTIVAGITAAAPDAOSDNFSGVAPDVTLSI
RQSSKRAPDPSSTGVGDVTHAKAVRTAADLGASVINISSTACVPAARAPDRL
GAALAYDVKNVIAAAGNTGAACPPQAQGVTRDSTVTVASPAWYDDYVLTVCS
VNAQEPSAFTLAGPWWDVAAATGEAVTSLSPFGDVTNRLGGQHGSIPISTSTAAHP
VSLGAALIRAFRPTLTARQVMORIESTAHHPAGMDPLVGVGTVDALAAVSSDSIPQA
GTATSDPAPVAVPVRRTSTPGSPDRALHTAFAGAAICLLALMATATASRRRLRPRN
GIAGD"
misc_feature
1757..1789
/gene="Rv3449"
/note="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862..1894
/gene="Rv3449"
/note="PS00137 Serine proteases, subtilase family,
histidine active site"
2456..2488
/gene="Rv3449"
/note="PS00138 Serine proteases, subtilase family, serine
active site"
complement(2810..4222)
/gene="Rv3450c"
complement(2810..4222)
/gene="Rv3450c"
/note="Rv3450c. (MTCY13E12.03c), len: 470. Function:
unknown but similar to Rv3895c MTCY15F10.17 (FASTA score:
1.6e-27; 36.4% identity in 475 aa overlap); possible
membrane spanning region near N-terminus, and to
V14967|MLC628.16 Mycobacterium leprae cosmid H628; (481
aa) opt: 708 Z-score: 710.9 E(): 4.5e-32; 32.9%
identity in 480 aa overlap, also similar to Rv3869
(MTV027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:O06317"
/translation="MPSFATTLVHVSQYRFLRLIECALLFGDVCAATGALRARTSL
ALGCVLAVAMGCAAFVALLRQPSALGQAPIVMGRESGALYVRVDVWHPIVNLASAR
LIAATNANPQVSESLGHTKRGPIILGIPGAPQLIDQPLAGAESAWAICDSNGGSTT
VVGPAEDSSAQVLTAEOMILVATFSGSPTLYLGGRRVAVHAIAPVAVWAIRLQGRV
PHVQAQSLNAVPEAPRIATPRIHGGGRASVCLPGFLGVGVKRIITKASGDEYVYVLED
GVORIGOVAADLLRFGDSQGSVNPTVADVIRVAPIVNTLTVSAFIPKPTPVGKSP
GRAVTLCTVMTFPAQGAARVAFLAGSGPPVPLGGVPTVLAGADGKGALLAVYLPPG
RSAYVAARSLSGGGTGTRYLVTDGVRFAIHDDDDVAHDGLGLPTAAIPAPVPLATLPS
GPGLSRANASVARDTVAPCP"
complement(4049..4081)
/gene="Rv3450c"
/note="PS00013 prokaryotic membrane lipoprotein lipid
attachment site"
complement(4228..4231)
/note="possible RBS, GGAG, for Rv3450c"
4388..5131
/gene="Rv3451"
4388..5131
/gene="Rv3451"
/note="Rv3451. (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MYCTU_Q50664
probable cutinase cy339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1

```

```

/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/db_xref="SWISS-PROT:O06318"
/translation="MGAGALITAVVLLIALGAVTVAFAADGCPDAFVTFARCTGSP
GICRGOAFVDSLRQQTGMHIGVYPVYAAASRLQLHGGDCANAIASHIKSMASCCPNT
KIYLVGSGQATVVDIVAGVPLGSI1SGSPPLPAAYADNAAVAFGNPSNRAGCSLSS
LSPLFGSKAIDIA/NPTDPTI/HVGGNEFSGHIDGYIPTTYTQAASFVOKIRAGSVPH
LPGSVPLQSGVLMQPGTAPAPESLHGR"
4706..4723
/misc_feature
/gene="Rv3451"
/note="PS00155 Cutinase, serine active site, GYSGQ"
5165..5168
/note="possible RBS, GAGG, for Rv3452"
5178..5858
/gene="Rv3452"
5178..5858
/gene="Rv3452"
/note="Rv3452. (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,
Query Match 100.0%; Score 18; DB 1; Length 43401;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 catgccgcgcctatctacty 18
|||||
Db 15872 CATGCCGTCGTATTGCTG 15855

RESULT 15
AF176573 9600 bp RNA VRL 18-AUG-1999
LOCUS Hepatitis C virus polyprotein precursor, gene, complete cds.
DEFINITION AF176573
ACCESSION AF176573
VERSION AF176573.1 GI:5738246
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus.
REFERENCE 1 (bases 1 to 9600)
AUTHORS Mokhonov,V.V., Samokhvalov,E.I., Novikov,D.V., Shatalov,A.G. and
Prilipov,A.G.
TITLE Molecular cloning HCV Russian isolate 1b from the serum of patient
with acute hepatitis
JOURNAL unpublished
REFERENCE 2 (bases 1 to 9600)
AUTHORS Mokhonov,V.V., Samokhvalov,E.I., Novikov,D.V., Shatalov,A.G. and
Prilipov,A.G.
TITLE Direct Submision
JOURNAL Submitted (09-AUG-1999) Molecular Genetics, Ivanovsky Virology
Institute, Gamaleya Str., 6, Moscow 123098, Russia
FEATURES
location/Qualifiers
1..9600
/organism="Hepatitis C virus"
/strain="274933R0"
/isolate="1b"
/db_xref="taxon:11103"
/country="Russia"
/note="isolated from acute hepatitis patient serum"
1..341
342..9374
/note="cleaved into C, E1, E2, p7, NS2, NS3, NS4a, NS4b,
NS5a, and NS5b proteins"
/codon_start=1
/product="polyprotein precursor"
/protein_id="AAD50312.1"
/db_xref="GI:5738247"
/translation="MSTNPKPQKTKRNTNRPPQDKFPGGQIVGGVYLLPRRGRPL
GVKATRKTSERSQPRQPIPKARHPEGRWAQPGYWPPLYNPGIGWAGWIIISPRG
SRPSWGPDPTRRRSRNLKVIDTLTCGFADLMGYIPLVGAPLUGVAKALAHGVRLVD

```

GVNATGNLPCCSFISFLLALLSCLTIPASAYEVNRNVGYVHVHTNDCSNSSIVYEAD
VIMHTPGVCPCVDGNTSCRWVALTPTLAARNASVPIAIRRHVDLLVGRARCSAMV
VGLCGSVFVPSQDLFTSPRRHQTVQDCNCSIYPGHISGHRMAWMMNWSPTAALVY
SQLLRIPQAIYDMVAGAHGVLAYISVGNNAKYMIVILLFAGVDGTHTTGGAA
ARATGFTSFPSLGSQKIQIOLINTNGSHINRTALNCNMISLQTCFLAALFYTYRFNAS
GCPERMASCRPIDKFDGOWGPIITVAEPDSSDORPCWYAPRPGCI VPAISOVCGPVYC
ETPSPVVYTTDRGCVPTWTGENETDVLNNTRPPLGNWFGCTWNNSTGFTKTCGG
PPNIGGAGNTLLTCPTDCFRKHPEATYKCGSGPMLTPRCLVDYPIRLMHPYCAVNF
TIFKRYMYGVGHRNAACNWTGERKCDLEDRDSELSPLLLSTTEWQVLPSCFTLL
PALSTGILTHQNTVDQYLYIGISAVIPFAIKWEYVLLFLLIADARVACCLMMMLL
IAQAEALLENLWNAASVAGAHGILSFLVFFCAAMY IKGRVPGAAYALYGVNPLLL
LLLAALPRAYAMDREMAACGAVFVGLVLLTSLPHYKVFELARI.IIMSOYFTTRAEAL
LOVWVPLNII RCGRDAIILMCVHSEL.IFDITKFLAILILGPLMMFOARITRVPYFVR
AQGLLRACMIYRKAAGHYIQMALMKLAALSYYVYDHLTPLODWAHAGLRDLAAVE
PVFSDMETKITWGAADTAACGDIISGLPVSARRGREILLGPADGLEEQGWRLLAPIT
PALSTGILTHQNTVDQYLYIGISAVIPFAIKWEYVLLFLLIADARVACCLMMMLL
AYSQTRGLLGCIIITSLTRDKNQVEGEVQVSTATQSFATCVCNGVCWTYVHAGTK
TLAGPKGTITQMYTNDQDLVQWQAPGARSLTPTCTCGSSDLXLYLVTRHADVIVRRRG
DSRGLSPRPISYLSKSSGGPLPCPSGHAVGIFRAAVCTRGVAKAVDFPVESMETT
MRSPVFTDSSPPAVPOTFOVAHLHAPTGGSKTKVPAAYAAQCKVILVLPNSVAATL
GFGAYMSKAGVDNLRGTGVRTITTGAPITYSTYKFLADGGCSGGAYDIIICDECHS
TDSITLIGITVLDQATAGARLVVLTATPPGSTVTPHPNIEIALSNTGEIPEFYK
AIPIEIKGGRHLIFCHSKKCKDELAALSGILGNVAYYHGLDVSVIPTSGNVVVA
TDALMTGTGTFDSVINDCTCTQTFVSLDPTFTIETTTVPQDAVSRRORRGTRHG
RGIYRVFTPGERPSGMEFSSVLCCEYDAGCAMELYTPAETSVRRLRAYLNTPLGPFVQ
DHLRFESVFTGLTHIDAHFLSOTKOAGDNFPYLVAYOATVCAQAAPPPSWDMWKC
LIRLPTLHGPTLLYRLGAVONETTLTHPTKYIMACMSADLEVTSTWLVGGLAE
ALAACTLTGSSVIVGRIVLSGKPAIIPDREVLQEFDEMEECASHLPYIBQGMQLAE
QFKKALGLLOATQKAAAPVPSWRALEAFWAKHIMNFISGVQYLAGSLTLPGN
PAIASLMAFTASITSPLTQYTLFNILGNWAAQALAPPASAASFVAGAGIAGAAGVSI
GLGKVLVDILAGYAGVAGALVAFKVMSEMPSTEDLVNLLPAIILSPGALYGVVCAA
ILRRHVGPGEAVOMNRLIAFASGNHVSPTHYVPESDAAARVQIILSNLTITQLLK
RLHOMINDECSPTCSGSLRDVMDWICVTLDFKTLRSKLLPRLPGVPFLSCQGYK
GWRGDIWOTPCGCAOIAGHVKNRMRIVGPRCTSNTHGTFP INAYTTGCTPSP
APNYSRALWRVAEEYVEITRVGDEHYVTGMTDNYKPCQCPAPAPEFTEVDGVRLHR
YAPVCKPLLDHEDTVFQGLNQYLVGSLQPEPEPDVAVLTSMLTDPSHITAEAKRRL
ARGSSQLASSASQLSAPSKATCTTRHSDPADLIEANLILRQEMGNGNITRVESEN
KVYIILDSFDPLRAEEDEREVSYPABILRKRKTFPSAMP IWARPDYNPPLLESWKDPDY
VPPVHGGCPLPPTKVPIPPRRKRVTVLTESTVSSALAEIATKTFGSSESAYDSGT
AAASPDQSDNGDSDVESYSSMPLEGEPCDPLSDGSMSTVSEASEDVCCSMS
YTWGALITPCGAESQLPINALNSLLRHLNLYATTSRASOROKKVTDFRLQVLD
DHYQDLKEMKAKASTYKAKLLSVEEACKLTPHSKSKFCYGAQKDVRLSSKAVNHI
HSWKDLLEDTEPIDTTIMAKNEVFCVQPEKGGKPKARLIVFPDLYRVCEKALYD
VYSTLPQAVMSAYGFQYSPQORVDFLVNAWKKKCPMGFSYDTRCFDSTVTESDIRV
EESIYQCCDLAPAEARQAIIRSLTERLYVGGPLTNSKGONCYRRCRASCGLVTTSCNTL
TCYLKATAACRAAKLRDCTLLVNGDDLVIICESAGTQEDAAASLRVFTeamTRYSApg
DLPOPEYDELITTSRNSNVSAHDASGRVYVLTDRDPTTLARAWEAARHTPVNSWL
GNIMAYPTLWARMILMTHFESILLFQLEKALDCOIYGAYYSIEPLDLPOIIORLH
GLSASFHSYSFGPEINRVASCLRLKICVPLRAWHRANVRAKLILSQGGRATCCCKYL
FNWAVRTKIKLTPIPAASQIDLSGMFVAGYGGDIYHSI.SKARPKWPMILCLLLLFVGV
GIYLLPNR*

3'UTR 1917 a 2887 c 2696 g 2100 L
BASE COUNT 1917 a 2887 c 2696 g 2100 L
ORIGIN

Query Match 91.1%; Score 16.4; DB 14; Length 9600;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 catgcccgtatgctg 18
|||||
Db 1125 CATGACGCTGATGCTG 1108

Search completed: April 2, 2002, 22:46:14
Job time: 12544 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nuclei: - nucleic search, using sw model

Run on: April 3, 2002, 08:04:13 ; Search time 165.4 seconds
(without alignments)
24.647 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaagtgattcgcg 18
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	18	100.0	12412	1	US-08-390-878-18
2	16.4	91.1	1535	3	US-08-464-052-1
3	16.4	91.1	1535	4	US-08-461-002-1
4	16.4	91.1	1535	4	US-08-689-411-1
5	16.4	91.1	1535	5	PCT-US94-09863-1
6	16.4	91.1	4411529	4	US-09-103-840A-1
7	14.4	80.0	573	2	US-08-709-912-18
8	14.4	80.0	573	2	US-09-047-370-18
9	14.4	77.8	1959	4	US-09-201-641-1
10	13.8	76.7	358	1	US-08-322-742-19
11	13.8	76.7	2321	4	US-09-198-839-2
12	13.8	76.7	2337	4	US-09-198-839-3
13	13.4	74.4	33	1	US-08-345-756-6
14	13.4	74.4	33	1	US-08-625-198-6
15	13.4	74.4	579	5	PCT-US96-05320A-1664
16	13.4	74.4	859	1	US-08-345-756-8
17	13.4	74.4	859	1	US-08-625-198-8
18	13.4	74.4	1902	3	US-09-041-991A-7
19	13.4	74.4	2039	1	US-08-345-756-5
20	13.4	74.4	2039	1	US-08-625-198-5
21	13.2	74.4	5467	2	US-08-605-106-7
22	13.2	73.3	30	1	US-08-547-182-4
23	13.2	73.3	60	3	US-08-577-121-3
24	13.2	73.3	60	4	US-08-585-700-3
25	13.2	73.3	60	4	US-09-149-161-3
26	13.2	73.3	60	5	PCT-US95-16916-3
27	13.2	73.3	1001	3	US-08-096-044C-2

28	13.2	73.3	1112	3	US-08-096-044C-9	Sequence 9, Appli
29	13.2	73.3	1530	1	US-08-149-105-1	Sequence 1, Appli
30	13.2	73.3	1530	1	US-08-317-847-1	Sequence 1, Appli
31	13.2	73.3	1557	5	PCT-US91-01327-11	Sequence 11, Appli
32	13.2	73.3	1557	6	5248670-1	Patent No. 5248670
33	13.2	73.3	1882	1	US-08-431-080-15	Sequence 15, Appli
34	13.2	73.3	1882	1	US-08-441-080-29	Sequence 29, Appli
35	13.2	73.3	1882	2	US-08-938-534-15	Sequence 15, Appli
36	13.2	73.3	1882	2	US-08-938-534-29	Sequence 29, Appli
37	13.2	73.3	4238	4	US-08-124-934A-5	Sequence 5, Appli
38	13.2	73.3	4238	5	PCT-US94-10080-5	Sequence 5, Appli
39	13.2	73.3	5211	1	US-08-447-411-1	Sequence 1, Appli
40	13.2	73.3	4403765	4	US-09-104-840A-2	Sequence 2, Appli
41	13.2	73.3	4403765	4	US-09-104-840A-2	Sequence 2, Appli
42	13.2	73.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
43	13	72.2	39	3	US-08-641-873-17	Sequence 17, Appli
44	13	72.2	1742	4	US-09-099-676-2	Sequence 2, Appli
45	13	72.2	1742	4	US-09-565-910-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-390-878-18
; Sequence 18, Application US/08390878
; Patent No. 5700684
; GENERAL INFORMATION:
; APPLICANT: Stovel, Charles K.
; APPLICANT: Mahalinas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Townsend Klorie and Crew
; STREET: One Market Plaza, Stewart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390.878
; FILING DATE: 17 FEB 1995
; CLASSIFICATION: 15
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Jan
; REGISTRATION NUMBER: 48,498
; REFERENCE/BOOKET NUMBER: 15,71A 1/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/544/9600
; TELEFAX: 415/544/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 100.0%; Score 18; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 acatcaagtgattcgcg 18
DB 468 ACATCAAGTGATTGCG 485

Wed Apr 3 08:33:23 2002

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

Query Match 91.1%; Score 16.4; DB 4; Length 1535;
Best Local Similarity 94.4%; Pred. No. 2.6;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgcg 18
|| |||||
DB 841 ACGTCAAAAGTGATTCGCG 858

RESULT 4
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA MOLECULE FRAGMENT'S ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,411
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-689-411-1
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-1

Query Match 91.1%; Score 16.4; DB 3; Length 1535;
Best Local Similarity 94.4%; Pred. No. 2.6;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgcg 18
|| |||||
DB 841 ACGTCAAAAGTGATTCGCG 858

RESULT 3
US-08-461-002-1
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-1
```

```

; TOPOLOGY: Circular
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc - "Synthetic DNA"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pTVX(3-190)
US-08-709-912-18

Query Match      80.0%; Score 14.4; DB 1; Length 573;
Best Local Similarity 93.8%; Pred. No. 30;

```

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgcgc 17
Db 471 CATTAAGTGATTCGC 456

RESULT 8

US-09-047-370-18/c
; Sequence 18, Application US/09047370

; Patent No. 5866408

; GENERAL INFORMATION:

; APPLICANT: Sung Dr., Wing L

; APPLICANT: Yaguchi Dr., Makoto

; APPLICANT: Ishikawa Dr., Kazuhiko

; TITLE OF INVENTION: Modification of Xylanase to Improve

; TITLE OF INVENTION: Thermophilicity, Alkalophilicity and

; TITLE OF INVENTION: Thermostability

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto

; STREET: 277 Park Ave.

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10172-0194

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: -US/09/047,370

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/709,912

; FILING DATE: 09-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Olsen Mr, Warren E

; REGISTRATION NUMBER: 27290

; REFERENCE/DOCKET NUMBER: 1039,2000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-2400

; TELEFAX: (212) 758-2982

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 573 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Synthetic DNA"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; IMMEDIATE SOURCE:

; CLONE: pTVX(3-190)

; US-09-047-370-18

Query Match 80.0%; Score 14.4; DB 2; Length 573;

Best Local Similarity 93.8%; Pred. NO. 30;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgcgc 17

Db 471 CATTAAGTGATTCGC 456

RESULT 9

US-09-201-641-1

; Sequence 1, Application US/09201641A

; Patent No. 6232530

;

; GENERAL INFORMATION:

; APPLICANT: Cunningham Jr, Francis X

; APPLICANT: DellaPenna, Dean

; TITLE OF INVENTION: Method for Regulating Carotenoid Biosynthesis in

; TITLE OF INVENTION: Marigolds

; FILE REFERENCE: Quest 41-162

; CURRENT APPLICATION NUMBER: US/09/201,641A

; CURRENT FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1959

; TYPE: DNA

; ORGANISM: Tagetes erecta

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (304)...(1836)

; OTHER INFORMATION: beta-cyclase

; US-09-201-641-1

Query Match 77.8%; Score 14; DB 4; Length 1959;

Best Local Similarity 100.0%; Pred. NO. 61;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtgatgc 15

Db 897 catcaaaagtgatgc 910

RESULT 10

US-08-322-742-19

; Sequence 19, Application US/08322742

; Patent No. 5688641

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; COMPUTER: IBM PS/2 Model 50Z or 55SX

; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

; SOFTWARE: WordPerfect (Version 5.1)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/322,742

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/948,823

; FILING DATE: September 1, 1992

; APPLICATION NUMBER: 07/844,296

; FILING DATE: February 28, 1992

; APPLICATION NUMBER: 07/552,216

; FILING DATE: February 28, 1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 00530/048003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 542-5070

; TELEFAX: (617) 542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 358

; TYPE: nucleic acid

; STRANDEDNESS: double

```

;
; TOPOLOGY: linear
US-08-322-742-19

Query Match 76.7%; Score 13.8; DB 1; Length 358;
Best Local Similarity 88.2%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
    ||||| ||||| |||
Db 262 CATCACAGTGATTCGCG 278

RESULT 11
US-09-198-839-2
; Sequence 2, Application US/09198839
; Patent No. 6190908
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray S.
; TITLE OF INVENTION: MODULATION OF POLYPEPTIDE DISPLAY ON MODIFIED
; FILE OF INVENTION: FILAMENTOUS PHAGE
; FILE REFERENCE: SCR21115
; CURRENT APPLICATION NUMBER: US/09/198,839
; CURRENT FILING DATE: 1998-12-24
; PRIOR APPLICATION NUMBER: PCT/US99/18207
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 60/096,326
; PRIOR FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2321
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PORFES II
US-09-198-839-2

Query Match 76.7%; Score 13.8; DB 4; Length 2321;
Best Local Similarity 88.2%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
    ||||| ||||| |||
Db 2043 catcaaatgttcacg 2059

RESULT 12
US-09-198-839-3
; Sequence 3, Application US/09198839
; Patent No. 6190908
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray S.
; TITLE OF INVENTION: MODULATION OF POLYPEPTIDE DISPLAY ON MODIFIED
; FILE OF INVENTION: FILAMENTOUS PHAGE
; FILE REFERENCE: SCR21115
; CURRENT APPLICATION NUMBER: US/09/198,839
; CURRENT FILING DATE: 1998-12-24
; PRIOR APPLICATION NUMBER: PCT/US99/18207
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 60/096,326
; PRIOR FILING DATE: 1998-08-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2337
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PORFES IV
US-09-198-839-3
```

```

Query Match 76.7%; Score 13.8; DB 4; Length 2337;
Best Local Similarity 88.2%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 18
    ||||| ||||| |||
Db 2059 catcaaatgttcacg 2075

RESULT 13
US-08-345-756-6
; Sequence 6, Application US/08345756
; Patent No. 5633438
; GENERAL INFORMATION:
; APPLICANT: Hasczynski, Chris
; APPLICANT: Harbour, Eric
; APPLICANT: Hattori, Jiro
; APPLICANT: Miki, Brian
; TITLE OF INVENTION: MICRUSPORE-SPECIFIC REGULATORY ELEMENT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,756
; FILING DATE: 22-NOV-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, STEPHEN A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/POCKET NUMBER: 33229/236/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
US-08-345-756 6

Query Match 74.4%; Score 13.4; DB 1; Length 33;
Best Local Similarity 93.3%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgttcgcg 16
    ||||| ||||| |||
Db 12 CATCAAACTGATGCG 26

RESULT 14
US-08-625-198-6
; Sequence 6, Application US/08625198
; Patent No. 5756324
; GENERAL INFORMATION:
; APPLICANT: Hasczynski, Chris
; APPLICANT: Harbour, Eric
; APPLICANT: Hattori, Jiro
; APPLICANT: Miki, Brian
```

;; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/625,198
;; FILING DATE: 01-APR-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/345,756
;; FILING DATE: 22-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, STEPHEN A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 33229/236/PIH1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 33 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-625-198-6

Query Match 74.4%; Score 13.4; DB 1; Length 33;
Best Local Similarity 93.3%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catcaaaagtgtatcg 16
|||||
Db 12 CATCAAAAGTGATCG 26

RESULT 15
PCT-US96-05320A-1664
; Sequence 1664, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US96/05320A
;; FILING DATE: April 22, 1996
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/476,102
;; FILING DATE: June 7, 1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/487,429
;; FILING DATE: June 7, 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Eric K. Stefie
;; REGISTRATION NUMBER: 36,688
;; REFERENCE/DOCKET NUMBER: 1488.014PC01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 1664:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 579 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: Linear
PCT-US96-05320A-1664

Query Match 74.4%; Score 13.4; DB 5; Length 579;
Best Local Similarity 93.3%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tcacaaatgattatcg 18
|||||
Db 446 TCACAAATGATTATCG 460

Search completed: April 3, 2002, 08:07:11
Job time: 42191 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 21:47:57 ; Search time 3835.7 Seconds
(without alignments)
50.427 Million cell updates/sec

Title: US-09-785-904-3
Perfect score: 18
Sequence: 1 acatcaagtgattcgcg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estfun:*
 - 2: em_esthum:*
 - 3: em_estirr:*
 - 4: em_estom:*
 - 5: em_estpl:*
 - 6: em_estba:*
 - 7: em_estro:*
 - 8: em_estov:*
 - 9: em_htc:*
 - 10: gb_est1:*
 - 11: gb_est2:*
 - 12: gb_htc:*
 - 13: gb_gss:*
 - 14: em_gss_fun:*
 - 15: em_gss_hum:*
 - 16: em_gss_inv:*
 - 17: em_gss_pln:*
 - 18: em_gss_pro:*
 - 19: em_gss_rtd:*
 - 20: em_gss_vrt:*
 - 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	16	88.9	135	10	AA249381	AA249381 j3441 seq
2	16	88.9	160	11	BF351466	BF351466 OVI-HT051
3	16	88.9	356	11	BF376314	BF376314 CM0-TN003
4	16	88.9	375	11	BF367020	BF367020 PM0-GN001
5	16	88.9	407	11	BF367003	BF367003 PM0-GN001
6	16	88.9	427	11	D44620	D44620 HUMSUPY017
7	16	88.9	415	11	BF772419	BF772419 IL5-IT002
8	16	88.9	594	10	AW750028	AW750028 PM2-BT054
9	15.4	85.6	166	11	BF944885	BF944885 CM1-NN019
10	15.4	85.6	238	10	AW248379	AW248379 2820521.5
11	15.4	85.6	269	11	F05463	F05463 HSCOBH041 n
12	15.4	85.6	301	11	F12193	F12193 HSC36H081 n

13	15.4	85.6	345	11	H14295	H14295 ym63b06.r1
14	15.4	85.6	360	11	BF947972	BF947972 CM1-NN019
15	15.4	85.6	380	11	BG639721	BG639721 SD04846.3
16	15.4	85.6	382	11	BG639722	BG639722 SD04865.3
17	15.4	85.6	383	11	BF947969	BF947969 CM1-NN019
18	15.4	85.6	385	10	AL134134	AL134134 DKF2P547J
19	15.4	85.6	414	11	BF944889	BF944889 CM1-NN019
20	15.4	85.6	415	11	BF927805	BF927805 IL5-IT022
21	15.4	85.6	416	11	BF944887	BF944887 CM1-NN019
22	15.4	85.6	467	11	BF490411	BF490411 AT26869.5
23	15.4	85.6	474	13	AQ692029	AQ692029 HS_5408_A
24	15.4	85.6	499	10	AA141880	AA141880 CK02646.3
25	15.4	85.6	508	10	AW390923	AW390923 MR3-ST020
26	15.4	85.6	524	13	AQ928124	AQ928124 KPC1-2-2
27	15.4	85.6	585	10	AL543240	AL543240 AL543240
28	15.4	85.6	606	11	BF238039	BF238039 601811750
29	15.4	85.6	610	10	AI513102	AI513102 CH01393.3
30	15.4	85.6	631	11	BG718674	BG718674 602696791
31	15.4	85.6	656	11	BF347028	BF347028 602021824
32	15.4	85.6	854	10	AL520879	AL520879 AI520879
33	15.4	85.6	885	10	AL559329	AL559329 AI559329
34	15.4	85.6	886	10	BF735986	BF735986 601405358
35	15.4	85.6	897	10	AL556646	AL556646 AL556646
36	15.4	85.6	916	11	BI115704	BI115704 602866051
37	15.4	85.6	940	10	AL536089	AL536089 AL536089
38	15.4	85.6	962	13	CNS02RLU	AL210747 Tetraodon
39	15.4	85.6	968	10	AL525534	AL525534 AL525534
40	15.4	85.6	972	10	AL551791	AL551791 AL551791
41	15.4	85.6	1094	11	W18183	W18183 IMAGE:20087
42	15	83.3	227	10	BE240803	BE240803 EST404852
43	15	83.3	237	11	BG382488	BG382488 298399 MA
44	15	83.3	267	10	AW774157	AW774157 EST333240
45	15	83.3	295	13	AQ091462	AQ091462 HS_3016_B

ALIGNMENTS

RESULT 1

AA249381

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA249381 135 bp mRNA EST 11-MAR-1997
j3441.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
5', mRNA sequence.
AA249381
AA249381.1 GI:1880285
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145)
Liew,C.C.
CDNAs from human fetal heart (1997)
Unpublished (1997)
Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewccc@utoronto.ca
PCR Primers
FORWARD: 5' GCCAAGCTCGAATTAACCCCTCACTAAAGGG 3'
BACKWARD: 5' CCACGTGATTCGTAATAGCTCACTATAGGCG 3'
Seq primer: 5' GAATTAACCCCTCACTAAAGGG 3'.
Location/Qualifiers
1. .135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site1: EcoRI; Site2:
XhoI; mRNA was purified from human fetal hearts (8-10

FEATURES

source

weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 36 a 43 c 33 g 23 t
 ORIGIN

Query Match 88.9%; Score 16; DB 10; Length 135;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
 |||||||

Db 28 CATCAAGTGCATCGC 43

RESULT 2
 BF351466 160 bp mRNA EST 22-NOV-2000
 LOCUS
 DEFINITION QV1-HT0517-020400-142-a08 HT0517 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF351466
 VERSION BF351466.1 GI:11310540
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 160)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

TITLE
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL.
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV1-HT0517-
 020400-142-a08&t3=2000-04-02&t4=1)
 Seq primer: puc 18 forward.
 Location/Qualifiers
 1..160
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0517"
 /dev_stage="Adult"
 /note="Organ: head_neck; Vector: puc18; Site:1: Smal;
 Site:2: Smal; A mini-library was made by cloning products
 derived from ORESTES PCR (O.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 46 a 48 c 40 g 26 t
 ORIGIN

Query Match 88.9%; Score 16; DB 11; Length 160;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
 |||||||

Db 20 CATCAAGTGCATCGC 35

RESULT 3

BF376314 356 bp mRNA EST 24-NOV-2000
 LOCUS
 DEFINITION CM0-TN0039-210800-507-c05 TN0039 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF376314
 VERSION BF376314.1 GI:11338339
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 356)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

REFERENCE

AUTHORS

TITLE

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL.
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1-CM0&t2-CM0-TN0039-
 210800-507-c05&t3=2000-08-21&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 59
 High quality sequence stop: 356.
 Location/Qualifiers
 1..356
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="TN0039"
 /dev_stage="Adult"
 /note="Organ: testis normal; Vector: puc18; Site:1: Smal;
 Site:2: Smal; A mini library was made by cloning products
 derived from ORESTES PCR (O.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 89 a 95 c 106 g 66 t
 ORIGIN

Query Match 88.9%; Score 16; DB 11; Length 356;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaaatgattgcgc 17
 |||||||

Db 74 CATCAAGTGCATCGC 89

RESULT 4

BF367020 375 bp mRNA EST 24-NOV-2000
 LOCUS
 DEFINITION PM0-GNU018-130900-003-e09 GNU018 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF367020

```

VERSION      BF367020.1  GI:11329071
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 375)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&l2=PM0-GN0018-
             130900-003-e09&t3=2000-09-13&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 375.
             Location/Qualifiers
             1..375-
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="GN0018"
             /dev_stage="Adult"
             /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
             ; Site_2: SmaI; A mini-library was made by cloning
             products derived from ORESTES PCR (U.S. Letters Patent
             application No. 196,716 - Ludwig Institute for Cancer
             Research) profiles into the puc 18 vector. Reverse
             transcription of tissue mRNA and cDNA amplification were
             performed under low stringency conditions."
             87 a 98 c 106 g 84 t

BASE COUNT   87 a 98 c 106 g 84 t
ORIGIN
Query Match      88.9%; Score 16; DB 11; Length 375,
Best Local Similarity 100.0%; Pred. NO. 4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaagtgatgcg 17
|||||
Db 311 CATCAAGTGATGCG 326

RESULT 5
LOCUS      BF367003 407 bp mRNA EST 24-NOV-2000
DEFINITION PM0-GN0018-040900-002-A08 GN0018 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF367003
VERSION     BF367003.1 GI:11329054
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 407)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&l2=PM0-GN0018-
             040900-002-A08&t3=2000-09-04&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 76.
             Location/Qualifiers
             1..407
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="GN0018"
             /dev_stage="Adult"
             /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
             ; Site_2: SmaI; A mini-library was made by cloning
             products derived from ORESTES PCR (U.S. Letters Patent
             application No. 196,716 - Ludwig Institute for Cancer
             Research) profiles into the puc 18 vector. Reverse
             transcription of tissue mRNA and cDNA amplification were
             performed under low stringency conditions."
             95 a 108 c 110 g 93 t 1 others

BASE COUNT   95 a 108 c 110 g 93 t 1 others
ORIGIN
Query Match      88.9%; Score 16; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. NO. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcaagtgatgcg 17
|||||
Db 329 CATCAAGTGATGCG 344

RESULT 6
LOCUS      D44620 415 bp mRNA EST 20-FEB-1998
DEFINITION HUMSHIP017 human brain cDNA Homo sapiens cDNA clone 003, mRNA
sequence.
ACCESSION  D44620
VERSION     D44620.1 GI:1572095
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 415)
AUTHORS      Hadano,S., Ishida,Y., Tomiyasu,H., Yamamoto,K., Bates,G.P. and
             Ikeda,J.
TITLE        Transcript map of the human chromosome 4p16.3 consisting of 627
             cDNA clones derived from 1 Mb of the Huntington's disease locus
             DNA Res. 3 (4), 239-255 (1996)
JOURNAL      97101646
MEDLINE      97101646
COMMENT      Contact: Shinji Hadano
             Japan Science and Technology Corporation, NeuroGenes Project, ICORP
             Univ. of Tokai School of Med.
             Boiseidai Isehara, Kanagawa 259-1193, Japan
             Tel: 81-463-91-5095
             Fax: 81-463-91-4993
             Email: shinji@nga.med.u-tokai.ac.jp.
```

```

FEATURES
Source
    Location/Qualifiers
    1. 415
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="003"
    /clone_lib="Human brain cDNA"
    /tissue_type="brain"
BASE COUNT      81 a 109 c 110 g 114 t 1 others
ORIGIN

Query Match      88.9%; Score 16; DB 11; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattccg 17
    |||||||
Db 360 CATCAAAAGTGATTCGC 375

RESULT 7
BF772419
LOCUS      BF772419      427 bp      mRNA      12-JAN-2001
DEFINITION IL5-IT0027-121200-322-b06 IT0027 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF772419
VERSION    BF772419.1 GI:12120319
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 427)
AUTHORS  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
    20202663
    Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=115&t2=115-IT0027-
    121200-322-b06&t3=2000-12-12&t4=1)
    Seq primer: puc 18 forward
    High quality sequence stop: 400.
    Location/Qualifiers
    1. 427
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="IT0027"
    /dev_stage="Adult"
    /note="Organ: epid_tumor; Vector: puc18; Site_1: Sma1;
    Site_2: Sma1; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the puc 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT      120 a 113 c 120 g 71 t 3 others
ORIGIN

FEATURES
Source
    Location/Qualifiers
    1. 427
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="IT0027"
    /dev_stage="Adult"
    /note="Organ: epid_tumor; Vector: puc18; Site_1: Sma1;
    Site_2: Sma1; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the puc 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
BASE COUNT      120 a 113 c 120 g 71 t 3 others
ORIGIN

Query Match      88.9%; Score 16; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattccg 17
    |||||||
Db 432 CATCAAAAGTGATTCGC 417

RESULT 9
AW750028/c
LOCUS      AW750028      594 bp      mRNA      28-APR-2000
DEFINITION PM2-BT0546-281299-003-f07 BT0546 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW750028
VERSION    AW750028.1 GI:7664960
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 594)
AUTHORS  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
    Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
    Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
    Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
    ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
    Simpson,A.J.
    Shotgun sequencing of the human transcriptome with ORF expressed
    sequence tags
    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
    20202663
    Contact: Simpson A.J.G.
    Laboratory of Cancer Genetics
    Ludwig Institute for Cancer Research
    Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
    Brazil
    Tel: +55-11-2704922
    Fax: +55-11-2707001
    Email: asimpson@ludwig.org.br
    This sequence was derived from the FAPESP/LICR Human Cancer Genome
    Project. This entry can be seen in the following URL
    (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-BT0546-
    281299-003-f07&t3=1999-12-28&t4=1)
    Seq primer: puc 18 forward
    High quality sequence stop: 581.
    Location/Qualifiers
    1. 594
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0546"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
    Sma1; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    ,716 - Ludwig Institute for Cancer Research) profiles
    into the puc 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
BASE COUNT      109 a 169 c 160 g 156 t
ORIGIN

Query Match      88.9%; Score 16; DB 10; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcaaaagtattccg 17
    |||||||
Db 432 CATCAAAAGTGATTCGC 417

RESULT 9

```

```

BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
LOCUS          CM1-NN0193-191000-484-g08 NN0193 Homo sapiens cDNA, mRNA sequence.
DEFINITION     BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
ACCESSION      BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
VERSION        BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
KEYWORDS       BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
SOURCE         BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
ORGANISM       BF944885      BF944885      166 bp      mRNA      EST      22-JAN-2001
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 166)
AUTHORS        Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE        20202663
COMMENT        Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&t2=CM1-NN0193-
               191000-484-g08&t3=2000-10-19&t4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 16
               High quality sequence stop: 166.
FEATURES       source
               1..166
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NN0193"
               /dev_stage="Adult"
               /note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
               Site_2: SmaI; A mini-library was made by cloning products
               derived from ORFEST PCR (U.S. Letters Patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the pUC 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."
BASE COUNT     41 a 44 c 50 g 31 t
ORIGIN
               1 acatcaaaagtgcgc 17
               |||||||
               Db 81 ACATCAAAAGTGATGGC 97

Query Match      85.6%; Score 15.4; DB 11; Length 166;
Best Local Similarity 94.1%; Pred. No. 7.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgcgc 17
   |||||||
Db 81 ACATCAAAAGTGATGGC 97

RESULT 10
LOCUS          AW248379      238 bp      mRNA      EST      07-JAN-2000
DEFINITION     AW248379      238 bp      mRNA      EST      07-JAN-2000
ACCESSION      AW248379      238 bp      mRNA      EST      07-JAN-2000
VERSION        AW248379      238 bp      mRNA      EST      07-JAN-2000
KEYWORDS       AW248379      238 bp      mRNA      EST      07-JAN-2000
SOURCE         AW248379      238 bp      mRNA      EST      07-JAN-2000
ORGANISM       AW248379      238 bp      mRNA      EST      07-JAN-2000
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 238)
AUTHORS        NTH-MGC http://mhc.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
               Other_ESTS: 2820521.3prime
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: DCTD/DPH cDNA Library Preparation: Ling
               Hong/Rubin laboratory cDNA Library Arrayed by: The I.M.A.G.E.
               Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing
               Project Clone distribution: MGC clone distribution information can
               be found through the I.M.A.G.E. Consortium/LLNL at:
               www.bio.llnl.gov/obrp/image/image.html Base Calling / Quality
               Scores: PHRED from University of Washington Genome Center. Vector
               Trimming: cross_match from University of Washington Genome Center
               PHRAP suite. Poly-T identification: patMatch.pl from Berkeley
               Drosophila Genome Project. University of Washington Genome Center:
               http://www.genome.washington.edu
               Plate: LICM4 row: G column: 18
               High quality sequence stop: 188.
FEATURES       source
               1..238
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:2820521"
               /clone_lib="NIH_MGC_7"
               /tissue_type="small cell carcinoma"
               /cell_line="MGC3"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
               EcoRI; cDNA made by oligo-dT priming. Directionally
               cloned into EcoRI/XhoI sites using the following 5'
               adaptor: GCCACGAG(G). Size-selected >500bp for average
               insert size 1.8kb. Library constructed by Ling Hong in
               the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT     53 a 68 c 86 g 31 t
ORIGIN
               1 acatcaaaagtgcgc 17
               |||||||
               Db 156 ACATCAAAAGTGATGGC 172

Query Match      85.6%; Score 15.4; DB 10; Length 238;
Best Local Similarity 94.1%; Pred. No. 7.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtgcgc 17
   |||||||
Db 156 ACATCAAAAGTGATGGC 172

RESULT 11
LOCUS          F05463      264 bp      mRNA      EST      19-FEB-1995
DEFINITION     F05463      264 bp      mRNA      EST      19-FEB-1995
ACCESSION      F05463      264 bp      mRNA      EST      19-FEB-1995
VERSION        F05463      264 bp      mRNA      EST      19-FEB-1995
KEYWORDS       F05463      264 bp      mRNA      EST      19-FEB-1995
SOURCE         F05463      264 bp      mRNA      EST      19-FEB-1995
ORGANISM       F05463      264 bp      mRNA      EST      19-FEB-1995
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 269)
AUTHORS        Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
               M.D., Duprat,S., Houlgatte,R., Juneau,M.N., Lamy,B., Lorenzo,F.,
               Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
               Sebastiani-Kabakchis,C. and Tessier,A.
TITLE          IMAGE: molecular integration of the analysis of the human genome
               and its expression
JOURNAL        C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE        95277534
COMMENT        Contact: Genethon

```

Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-0bh04
Seq primer: (-21)M13 universal.

FEATURES

source
1..269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-0bh04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 65 a 75 c 90 g 39 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 269;
Best Local Similarity 94.1%; Pred. No. 7.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgc 17
|||||
Db 137 ACATCAAAAGTGC 153

RESULT 12

F12193 301 bp mRNA EST 13-MAR-1995
LOCUS HSC36H081 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION C-36H08, mRNA sequence.
ACCESSION F12193
VERSION F12193.1 GI:706535
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houllgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouillot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-36h08
Seq primer: (-21)M13 universal.

FEATURES

source
1..301
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="c-36h08"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 73 a 78 c 99 g 47 t
ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 301;
Best Local Similarity 94.1%; Pred. No. 7.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acatcaaaagtgcgc 17
|||||
Db 135 ACATCAAAAGTGC 151

RESULT 13

H14295 345 bp mRNA EST 10-JUL-1995
LOCUS ym53b06.r1 Soares adult brain N2b4HB55Y Homo sapiens cDNA clone
DEFINITION IMAGE:163571 5' similar to SP:KG3A_RAT P18265 GLYCOGEN SYNTHASE
KINASE-3 ALPHA ;, mRNA sequence.
ACCESSION H14295
VERSION H14295.1 GI:879115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1920
High quality sequence stops: 179
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1920 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 179.

FEATURES

source
1..345
/organism="Homo sapiens"
/db_xref="GDB:583540"
/db_xref="taxon:9606"
/clone="IMAGE:163571"
/clone_lib="Soares adult brain N2b4HB55Y"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: brain; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAACTGAGTGGGAGCGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M.Fatima Bonaldo. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT 90 a 81 c 100 g 71 t 3 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 345;
Best Local Similarity 94.1%; Pred. No. 8.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtattgc 17

Db 73 ACATCAAAAGTATTGCC 89

RESULT 14

BF947972/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

LABORATORY

ADDRESS

TELEPHONE

FAX

EMAIL

PROJECT

URL

SEQUENCE

PRIMER

High quality sequence start: 35

High quality sequence stop: 360.

Location/Qualifiers

1. .360

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NN0193"

/dev_stage="Adult"

/note="organ: nervous_normal; Vector: puc18; Site_1: SmaI;

FEATURES

Source

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches 1; Indels

0; Gaps 0;

Qy 2 catcaaaagtattgc 18

Db 2 catcaaaagtattgc 18

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches 1; Indels

0; Gaps 0;

Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 79 a 99 c 91 g 91 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 360;
Best Local Similarity 94.1%; Pred. No. 8.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 acatcaaaagtattgc 17

Db 334 ACATCAAAAGTATTGCC 318

RESULT 15

BG639721/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

LABORATORY

ADDRESS

TELEPHONE

FAX

EMAIL

PROJECT

URL

SEQUENCE

PRIMER

High quality sequence start: 289.

High quality sequence stop: 289.

Location/Qualifiers

1. .380

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="SD04846"

/clone_lib="SD04846"

/lab_host="DH5-alpha"

/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized

fractionated cDNAs were directly ligated into pOT2.

Plasmid cDNA library."

BASE COUNT 103 a 70 c 83 g 124 t

ORIGIN

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches 1; Indels

0; Gaps 0;

Qy 2 catcaaaagtattgc 18

Db 2 catcaaaagtattgc 18

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches 1; Indels

0; Gaps 0;

Qy 2 catcaaaagtattgc 18

Db 2 catcaaaagtattgc 18

Query Match

Best Local Similarity

Matches 16; Conservative

0; Mismatches 1; Indels

0; Gaps 0;

Db 379 CATCAAGAGATCGCG 363
||||||| |||||||

Search completed: April 2, 2002, 21:48:01
Job time: 10826 sec

PS Claim 7; Page 45; 67pp; English.

XX A DNA molecule (AAW02303) codes for a protein (AAW02303) that confers
CC on Mycobacterium tuberculosis an ability to survive within
CC macrophages. It is a fragment of a larger molecule (AAW02303) obtd.
CC by ligating M. tuberculosis genomic DNA fragments into pBluescript
CC II vector and screening recombinant E. coli strains for HeLa cell-
CC invasive clones. It can be incorporated into a vector and used
CC for prodn. of recombinant macrophage survival protein, which is
CC useful in vaccines or for facilitating uptake of other materials,
CC e.g. therapeutic genes, into mammalian cells. A second DNA molecule
CC (AAW02302) codes for a protein (AAW02302) that confers on M.
CC tuberculosis the ability to enter mammalian cells.

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 650;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18

Db 189 CATGCCGTCGTATTGCTG 172

RESULT 2

AAV18649/c
ID AAV18649 standard; DNA; 650 BP.

XX AAV18649;

XX 03-JUL-1998 (first entry)

XX DNA for M. tuberculosis cellular uptake protein fragment.

XX Cellular uptake protein; vaccine; infection; ds.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 1..649

FT CDS /*tag= a

FT /*note= "stop codon not given"

XX WO9805784-A1.

XX 12-FEB-1998.

XX 06-AUG-1997; 97WO-US13056.

XX 07-AUG-1996; 96US-0689411.

XX (CONN-) CONNAUGHT LAB LTD.

PA (CORR) CORNELL RES FOUND INC.

XX Chong P, Riley LW;

XX WPI; 1998-145620/13.

DR P-PSDB; AAW47543.

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
PT therapeutic agents e.g. antibiotics, also useful in vaccines

PS Disclosure; Page 14; 82pp; English.

XX The present sequence encodes a Mycobacterium tuberculosis cellular
CC uptake protein fragment, which confers on M. tuberculosis an
CC ability to survive within macrophages.

CC The protein can be used in a vaccine to prevent M. tuberculosis

CC infection, and provide for the uptake in cells of, e.g.

CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies

CC raised against it can be used to treat mammals already exposed to

CC M. tuberculosis, to induce a passive immunity and prevent disease
CC occurrence.

XX Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 650;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18

Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3

AAQ89200/c

ID AAQ89200 standard; DNA; 1535 BP.

XX AAQ89200;

DT 09-OCT-1995 (first entry)

XX Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
DE protein.

XX Vaccine; tuberculosis; ss.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 1..1535

FT CDS /*tag= a

XX WO9506726-A.

XX 09-MAR-1995.

XX 01-SEP-1994; 94WO-US09863.

XX 02-SEP-1993; 93US-0118442.

XX (CORR) CORNELL RES FOUND INC.

XX Riley LW;

XX WPI; 1995-115442/15.

DR P-PSDB; AAR71931.

XX DNA encoding for cellular uptake of Mycobacterium tuberculosis -
PT used to develop prods for vaccines, passive immunisation and
PT diagnosis and cellular uptake of other materials

XX Claim 2; Page 9-11; 46pp; English.

XX The isolated DNA molecule of the invention confers on M.

CC tuberculosis an ability to enter cells and to survive within

CC macrophages. It encodes a polypeptide having a mol. wt. of about 50-

CC 55 kDa, pref. 52 kDa. The AA sequence represents a highly

CC hydrophilic protein with a hydrophobic region at its carboxy

CC terminus. It could be a secreted protein, a cytoplasmic protein, or

CC a surface protein with its carboxy terminus attached to the outer

CC membrane of the organism. The deduced AA sequence is in AAR71931.

XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 100.0%; Score 18; DB 16; Length 1535;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18

|||||

Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 4

AAAT33656/c
ID AAT33656 standard; DNA; 1535 BP.

XX AC AAT33656;

XX DT 22-NOV-1996 (first entry)

XX DE M. tuberculosis cellular uptake gene region.

XX KW Cellular uptake; cell entry; macrophage; passive immunisation;

XX KW vaccine; gene therapy; ds.

XX OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).

XX FH Key Location/Qualifiers

XX FT 181..810

XX FT /*tag= a

XX FT /label= ORF-1

XX FT /product= cell entry protein

XX FT 886..1535

XX FT /*tag= b

XX FT /label= ORF-2

XX FT /product= macrophage survival protein

XX PN W09626275-A1.

XX PD 29-AUG-1996.

XX PF 20-FEB-1996; 96WO-US02155.

XX PR 22-FEB-1995; 95US-0392210.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Riley LW;

XX DR WPI; 1996-425086/42.

XX DR P-PSDB; AAW02301.

XX PT DNA giving M. tuberculosis ability to enter mammalian cells -
PT and/or survive within macrophage(s), useful in vaccines to protect
PT mammals against Mycobacterium tuberculosis infection

XX PS Claim 2; Page 41; 67pp; English.

XX CC A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an
XX CC ability to enter mammalian cells and to survive within macrophages.
XX CC The encoded protein sequence is given in AAW02301. The DNA was obtd.
XX CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II
XX CC vector and screening recombinant E. coli strains for beta cell-
XX CC invasive clones. The DNA includes 2 separate coding regions (see
XX CC also AAT33657-58) coding for the cell entry (AAW02302) and macrophage
XX CC survival (AAW02303) proteins. It can be used to produce the cellular
XX CC uptake proteins used as vaccines or to facilitate uptake of other
XX CC materials, e.g. therapeutic genes.

XX SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 1535;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18

|||||

Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 5

AAV18647/c
ID AAV18647 standard; DNA; 1535 BP.

XX AC AAV18647;

XX DT 03-JUL-1998 (first entry)

XX DE RNA for M. tuberculosis cellular uptake protein fragment.

XX KW Cellular uptake protein; vaccine; infection; ds.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers

XX FT 1..1534

XX FT /*tag= a

XX FT /note= "stop codon not given"

XX PN W09805784-A1.

XX PD 12-FEB-1998.

XX PF 06-AUG-1997; 97WO-US13056.

XX PR 07-AUG-1996; 96US-0689411.

XX PA (CORR) CORNELL RES FOUND INC.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Chong P, Riley LW;

XX DR WPI; 1998-145620/13.

XX DR P-PSDB; AAW47541.

XX PT Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
therapeutic agents e.g. antibiotics, also useful in vaccines

XX PS Disclosure; Pages 9-10; 82pp; English.

XX CC The present sequence encodes a Mycobacterium tuberculosis cellular
XX CC uptake protein fragment, which confers on M. tuberculosis an
XX CC ability to enter mammalian cells and to survive within macrophages.
XX CC The protein can be used in a vaccine to prevent M. tuberculosis
XX CC infection, and provide for the uptake in cells of, e.g.
XX CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
XX CC raised against it can be used to treat mammals already exposed to
XX CC M. tuberculosis, to induce a passive immunity and prevent disease
XX CC occurrence.

XX SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 1535;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18

|||||

Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 6

AAAT33537/c
ID AAT33537 standard; DNA; 12412 BP.

XX AC AAT33537;

XX DT 15-FEB-1998 (first entry)

XX DE BCG deletion region 3 and flanking sequences.

XX KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;

XX KW mycobacteria; vaccine; infection; marker; ss.

```

XX OS Mycobacterium bovis strain BCG.
XX FH Key Location/Qualifiers
XX FT 1406..10673
XX FT misc_feature /*tag= a
XX FT /*note= "BCG delta 1 deletion region"
XX PN WO9625519-A1.
XX PD 22-AUG-1996.
XX PF 15-FEB-1996; 96WO-US01938.
XX PR 17-FEB-1995; 95US-0390878.
XX PA (PATH-) PATHOGENESIS CORP.
XX PI Mahairas GG, Stover CK;
XX WPI; 1996-393419/39.
XX Detecting markers for avirulence in Mycobacterium - used in
XX production of vaccines against bacterial infection, and to detect
XX bacterial infection
XX Example 1; Fig 3; 66pp; English.
XX This DNA sequence comprises Mycobacterium bovis BCG deletion
XX sequence BCGdelta3. A specific genetic deletion of this region
XX results in an avirulence phenotype of the mycobacterium. 2 Other
XX deletion regions (see AAT33535 and AAT33536) have also been detected.
XX Identification involved screening a BCG cosmid library with a
XX radiolabeled probe obtained following DNA subtraction between
XX virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
XX The deletions provide useful markers for the identification of an
XX avirulent, or a virulent, mycobacterial phenotype. Determination
XX of avirulence requires the detection of the presence or absence of
XX the deletion; the deletions are detected either by detecting the
XX presence or absence of deletion junctions (see AAT33338-46), or by
XX detecting the presence or absence of the sequences contained within
XX the deletion. Deletion polypeptides are used as components of
XX immunological assays and in vaccines.
XX SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;

Query Match 100.0%; Score 18; DB 17; Length 12412;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtctattgctg 18
Db |||||||
705 CATGCCGTCGTATGCTG 688

RESULT 7
AAZ53670
ID AAZ53670 standard; DNA; 1368 BP.
XX AC AAZ53670;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 311 partial DNA sequence SEQ ID NO:1289.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX antibacterial; gene therapy; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO9957280-A2.

```

```

XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarcellini M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX P-PSDB; AAY74908.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 7; Page 70y; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX PCR primers used in the exemplification of the present inventions. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX SQ Sequence 1368 BP; 247 A; 377 C; 448 G; 256 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 1468;
Best Local Similarity 94.1%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgcgcgtctattgctg 18
Db |||||||
305 atgcgcgtctattgctg 321

RESULT 8
AAZ12075
ID AAZ12075 standard; DNA; 1779 BP.
XX AC AAZ12075;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria gonorrhoeae complete ORF61 sequence.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhoea; ss.
XX OS Neisseria gonorrhoeae.
XX PN WO9924578-A2.
XX PD 20-MAY-1999.

```

```

XX PF 09-OCT-1998; 98WO-IB01665.
XX PR
XX PA
XX PA (CHIR-) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venturi JC;
XX XX WPI: 2000-062150/05.
XX DR P-PSDB; AAY74911.
XX XX
XX XX Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics
XX PT
XX PS Claim 7; Page 712-713; 1453pp; English.
XX XX
XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polypeptides can be used in the
XX CC manufacture of medicaments for treating or preventing infection due to
XX CC Neisseria meningitidis (e.g. meningitis and septicaemia), to detect the
XX CC presence of Neisseria meningitidis, or to raise antibodies. They may also
XX CC be used to screen for agonists or antagonists, which may themselves
XX CC have use as antibacterial agents. The polynucleotides of the invention
XX CC may also be used in gene therapy protocols.
XX XX
XX SQ Sequence 1779 BP; 367 A; 478 C; 599 G; 335 T; 0 other;

Query Match 85.6%; Score 15.4; DB 20; Length 1779;
Best Local Similarity 94.1%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgcgtattctg 18
||||| |||||||
DB 716 atgccgcgtattctg 732

RESULT 9
AAZ53673
ID AAZ53673 standard; DNA; 1779 BP.
XX AC
XX AC AAZ53673;
XX DT 21-MAR-2000 (first entry)
XX DE
XX DE Neisseria gonorrhoeae ORF 311 partial DNA sequence SEQ ID NO:1205.
XX XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy; ds.
XX XX
XX OS Neisseria gonorrhoeae.
XX XX
XX PN WO957280-A2.
XX XX
XX XX 11-NOV-1999.
XX PD
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.

XX PF 09-OCT-1998; 98WO-IB01665.
XX PR
XX PA
XX PA (CHIR-) CHIRON SPA.
XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX PI WPI: 1999-327407/27.
XX DR P-PSDB; AAY38618.
XX DR
XX XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX PT diagnosis, treatment and prevention of infection
XX PT
XX PS Claim 9; Page 175; 524pp; English.
XX XX
XX CC Nucleotide sequences AAZ11972-Z12358 represent open reading frames
XX CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
XX CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
XX CC fragments, their nucleic acids and antibodies are used for diagnosis,
XX CC prevention (as vaccines) or treatment of Neisseria meningitidis,
XX CC such as meningitis, septicaemia and gonorrhea. Both organisms
XX CC are closely related. Fragments of the nucleic acids are useful
XX CC as hybridisation probes and antisense reagents.
XX XX
XX SQ Sequence 1779 BP; 367 A; 478 C; 599 G; 335 T; 0 other;

Query Match 85.6%; Score 15.4; DB 20; Length 1779;
Best Local Similarity 94.1%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgcgtattctg 18
||||| |||||||
DB 716 atgccgcgtattctg 732

RESULT 10
AAZ14061
ID AAZ14061 standard; DNA; 5859 BP.
XX AC
XX AC AAZ14061;
XX DT 09-JUL-1996 (first entry)
XX DE
XX DE N. gonorrhoeae lgt locus.
XX XX
XX KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
XX KW vaccine; ds.
XX XX
XX OS Neisseria gonorrhoeae strain F62.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 445..1491
XX FT /*tag= a
XX FT /*product= LgtA
XX FT misc_feature 699..715
XX FT /*tag= b
XX FT /*note= "poly-G tract"
XX FT CDS 1491..2330
XX FT /*tag= c
XX FT /*product= LgtB
XX FT CDS 2342..3262
XX FT /*tag= d
XX FT /*product= LgtC
XX FT misc_feature 2499..2508
XX FT /*tag= e

```

```

FT CDS /note= "poly-G tract"
FT 3322..4335
FT /*tag= f
FT /product= LgtD
FT misc_feature 3576..3586
FT /*tag= g
FT /note= "poly-G tract"
FT 4354..5196
FT /*tag= h
FT /product= LgtE
FT
XX WO9610086-A1.
XX
XX
XX PD 04-APR-1996.
XX
XX PF 25-SEP-1995; 95WO-US12317.
XX
XX PR 26-SEP-1994; 94US-0312387.
XX
XX PA (UVRQ ) UIV ROCKEFELLER.
XX
XX Gotschlich EC;
XX
XX WPI; 1996-200924/20.
XX
XX DR P-PSDB; AAR91311, AAR91312, AAR91313, AAR91314, AAR91315.
XX
XX Nucleic acids encoding glycosyl transferase(s) - used in the
PT diagnosis of infection with Neisseria and for the biosynthesis of
PT oligo:saccharide(s)
XX
XX Claim 1; Fig 2g-m; 81pp; English.
XX
XX The lgt locus (AAT14061) of Neisseria gonorrhoeae F62 contains
CC 5 open reading frames, lgtA, lgtB, lgtC, lgtD and lgtE, coding for
CC 5 glycosyltransferases (see also AAR91311-15) involved in gonococcal
CC lipooligosaccharide (LOS) biosynthesis. The sequence was
CC constructed from 2 clones isolated from an F62 gene bank in
CC lambda-200 following screening with plasmid pRI0PI. 3 Of the
CC coding sequences contain poly-G tracts that make them susceptible
CC to premature termination. The lgt coding sequences can be used
CC for prodn. of recombinant Lgt glycosyltransferases that are utilised
CC in biosynthesis of LOS useful in vaccine prepn.
XX
XX Sequence 5859 BP; 1411 A; 1462 C; 1661 G; 1325 T; 0 other;
SQ
Query Match 85.6%; Score 15.4; DB 17; Length 5859;
Best Local Similarity 94.1%; Prev. No. 54;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 atgcgcgtcgattgctg 18
Db 2796 atgcgcgcgtattgctg 2812
||||| |||||||

RESULT 11
AAT49230
ID AAT49230 standard; DNA; 5859 BP.
XX
XX AC AAT49230;
XX
XX DT 21-MAR-1997 (first entry)
XX
XX Lipo-oligosaccharide (including polyglycosyltransferase) gene.
DE
DE Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KW N-acetylgalactosaminyl transferase; lipo-oligosaccharide; ss.
KW
XX Neisseria gonorrhoeae ATCC 33084.
OS
XX Key Location/Qualifiers
FH 1..381
CDS /*tag= a
FT

```

```

FT CDS 445..1491
FT /*tag= b
FT /product= polyglycosyltransferase
FT 2342..3262
FT /*tag= c
FT 3322..4335
FT /*tag= d
FT 4354..5196
FT /*tag= e
XX
XX WO9640971-A1.
XX
XX PD 19-DEC-1996.
XX
XX PF 03-JUN-1996; 96WJ-US08323.
XX
XX PR 07-JUN-1995; 95US-0478140.
XX
XX PA (NEOS-) NEOSE TECHNOLOGIES INC.
XX
XX Buczala SL, Johnson KF, Koth S;
XX
XX WPI; 1997-052451/05.
XX
XX DR P-PSDB; AAW06576, AAW06577, AAW06578, AAW06579, AAW06580.
XX
XX Transfer of at least 2 saccharide units using
PT polyglycosyltransferase - isolated from N. gonorrhoeae, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety
XX
XX Disclosure; Fig 2A-B; 38pp; English.
XX
XX A lipooligosaccharide-encoding gene region (AAT49230) of Neisseria
CC gonorrhoeae ATCC 33084 includes a coding sequence for a novel N-
CC polyglycosyltransferase (PGTase) (AAW06576) that catalyses the
CC addition of both GlcNAc and GalNAc disaccharides to a single
CC galactose moiety. The products (AAW06577-80) of the other coding
CC sequences of this region are not identified. The PGTase gene can
CC be obtd. using standard techniques and incorporated into a vector
CC to allow prodn. of the PGTase in transformed host cells. The
CC enzyme is useful in the synthesis of oligosaccharides.
XX
XX Sequence 5859 BP; 1407 A; 1462 C; 1661 G; 1329 T; 0 other;
SQ
Query Match 85.6%; Score 15.4; DB 18; Length 5859;
Best Local Similarity 94.1%; Prev. No. 54;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 atgcgcgtcgattgctg 18
Db 2796 atgcgcgcgtattgctg 2812
||||| |||||||

RESULT 12
AAA53978/G
ID AAA53978 standard; DNA; 27425 BP.
XX
XX AC AAA53978;
XX
XX DT 08-FEB-2001 (first entry)
XX
XX Dmp53 tumour suppressor gene (genomic DNA).
DE
DE p53; tumour suppressor gene; insect; phenotype; metazoan;
KW invertebrate; screening; pharmaceutical; pesticide; mis-expression;
KW mutation; modulation; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200055178-A1.
XX
XX PD 21-SEP-2000.

```

```

XX 13-MAR-2000; 2000WO-US06602.
XX PF
XX 16-MAR-1999; 99US-0268969.
XX PR
XX 23-FEB-2000; 2000US-0184373.
XX PR
XX (EXEL-) EXELIXIS INC.
XX PA
XX Buchman AR, Platt DM, Ollman MM, Young LM, Demsky MR, Keegan KP;
XX PI Friedman L, Kopczyński C, Larson JS, Robertson SA;
XX PI
XX WPI; 2000-638178/61.
XX DR
XX Novel p53 tumor suppressor gene encoding a protein useful for
XX PT genetically modifying metazoan invertebrate organisms, such as insects
XX PT for screening compounds of pharmaceutical use or a pesticide
XX PT
XX Example 7; Page 80-89; 98pp; English.
XX PS
XX Insect p53 tumour suppressor genes can be used to genetically modify
XX CC metazoan invertebrate organisms, such as insects and worms, or
XX CC cultured cells, resulting in p53 expression or mis-expression. The
XX CC tumour suppressor genes, a p53 polypeptide or genetically modified
XX CC organisms or cells are used in screening assays to identify compounds
XX CC or molecules, preferably a pharmaceutical agent or a pesticide, that
XX CC modulates p53 activity. The genetically modified organisms or cells
XX CC are also useful for studying p53 activity by detecting the phenotype
XX CC caused by the expression or mis-expression of the p53 protein in the
XX CC insect. The method additionally comprises observing a second insect
XX CC having the same genetic modification as the previous one, which
XX CC causes the expression or mis-expression of the p53 protein, where the
XX CC second animal additionally comprises a mutation in a desired gene and
XX CC differences between the phenotype of the first and second identifies
XX CC the desired gene as capable of modifying the function of the gene
XX CC encoding the p53 protein. The genetically modified organisms or
XX CC cells are also useful for identifying other genes modulating the
XX CC function of, or interaction with the p53 gene.
XX CC
XX Sequence 27425 BP; 7309 A; 6492 C; 6305 G; 7319 T; 0 other;

Query Match 85.6%; Score 15.4; DB 21; Length 27425;
Best Local Similarity 94.1%; Pred. No. 60;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18
Db | ||||| |||||
9553 AGGCCGTCGTATTGCTG 9537

RESULT 13
AAQ20926/c
ID AAQ20926 standard; DNA; 742 BP.
XX AC AAQ20926;
XX XX
XX 06-MAY-1992 (first entry)
XX DT
XX C10-E15 DNA fragment encoding NANBH-specific antigen polypeptide.
XX DE
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX KW
XX Non-A non-B hepatitis virus.
XX OS
XX Key Location/Qualifiers
XX FH 3..742
XX FT /*tag= a
XX PN EP468657-A.
XX PD 29-JAN-1992.
XX XX
XX 08-JUL-1991; 91EP-0306158.
XX DE
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX KW
XX Non-A non-B hepatitis virus.
XX OS
XX Key Location/Qualifiers
XX FH 3..742
XX FT /*tag= a
XX PN EP468657-A.
XX PD 29-JAN-1992.
XX XX
XX 08-JUL-1991; 91EP-0306158.

```

```

XX 20-DEC-1990; 90JP-0413844.
XX PR 09-JUL-1990; 90JP-0180889.
XX PR 30-NOV-1990; 90JP-0339589.
XX XX
XX (TOFU) TONEN CORP.
XX PA
XX Maki N, Yamaguchi K, Toyoshima A, Kohara M;
XX PI
XX WPI; 1992-034390/05.
XX DR P-PSDB; AAR20723.
XX DR
XX Non-A, non-B hepatitis-specific antigen polypeptide - for detection
XX PT of hepatitis virus gene or antibody directed against virus
XX PT
XX Claim 18; Fig 18; 78pp; English.
XX PS
XX The sequence is that of a DNA fragment encoding a non-A non-B (NANB)
XX CC hepatitis-specific antigen polypeptide, it can be used to produce the
XX CC polypeptide recombinantly. It is derived from the DNA clone C10-E15.
XX CC It can be used to give PCR primers which are capable of detecting NANB
XX CC hepatitis with high accuracy. See also AAQ20617-Q20629 and
XX CC AAQ20922-Q20925.
XX CC
XX Sequence 742 BP; 129 A; 217 C; 209 G; 187 T; 0 other;

Query Match 83.3%; Score 15; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgtattg 15
Db | ||||| |||||
372 CATGCCGTCGTATTG 358

RESULT 14
AAQ20923/c
ID AAQ20923 standard; DNA; 932 BP.
XX AC AAQ20923;
XX XX
XX 06-MAY-1992 (first entry)
XX DT
XX C10-E12 DNA fragment encoding NANBH-specific antigen polypeptide.
XX DE
XX Non-A non-B hepatitis virus; recombinant; detection; ss.
XX KW
XX Non-A non-B hepatitis virus.
XX OS
XX Key Location/Qualifiers
XX FH 3..942
XX FT /*tag= a
XX PN EP468657-A.
XX PD 29-JAN-1992.
XX XX
XX 08-JUL-1991; 91EP-0306158.
XX DE
XX 20-DEC-1990; 90JP-0413844.
XX PR 09-JUL-1990; 90JP-0180889.
XX PR 30-NOV-1990; 90JP-0339589.
XX XX
XX (TOFU) TONEN CORP.
XX PA
XX Maki N, Yamaguchi K, Toyoshima A, Kohara M;
XX PI
XX WPI; 1992-034390/05.
XX DR P-PSDB; AAR20720.
XX DR
XX Non-A, non-B hepatitis-specific antigen polypeptide - for detection
XX PT of hepatitis virus gene or antibody directed against virus

```

XX PS Claim 15; Fig 15; 78pp; English.

XX CC The sequence is that of a DNA fragment encoding a non-A non-B (NANB) hepatitis-specific antigen polypeptide, it can be used to produce the CC polypeptide recombinantly. It is derived from the DNA clone C10-E12. CC It can be used to give PCR primers which are capable of detecting NANB CC hepatitis with high accuracy. See also AAQ20617-Q20629 and CC AAQ20922-Q20926.

XX SQ Sequence 932 BP; 173 A; 276 C; 257 G; 226 T; 0 other;

Query Match 83.3%; Score 15; DB 13; Length 932;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgcgcgtcgattg 15
|||||

Db 438 CATGCCGTCGTATTG 424

RESULT 15

AAAX19484
ID AAX19484 standard; DNA; 1569 BP.

XX AC AAX19484;

XX DT 21-MAY-1999 (first entry)

XX DE Streptococcus pneumoniae ffh encoding DNA.

XX KW Streptococcus pneumoniae; ffh; fifty-four homologue; antibacterial;
infection; otitis media; conjunctivitis; bacteraemia; sinusitis;
pleural empyema; endocarditis; meningitis; ds.

XX OS Streptococcus pneumoniae.

XX PN EP900843-A2

XX PD 10-MAR-1999.

XX PF 20-AUG-1998; 98EP-0306685.

XX PR 02-SEP-1997; 97US-0923772.

XX PA (SMIK) SMITHLINE BEECHAM.

XX PI Black MT;

XX DR WPI; 1999-155936/14.

XX DR P-PSDB; AAW99658.

XX CC New Streptococcus pneumoniae Fifty-Four Homologue (Ffh) polypeptide
PT and polynucleotide - useful as diagnostic reagents and for
PT prevention and treatment of Streptococci infections, which cause
PT otitis media, sinusitis and conjunctivitis

XX PS Claim 4; Page 5-6; 21pp; English.

XX CC The present sequence encodes the Streptococcus pneumoniae fifty-four
homologue (ffh) protein, which is a component of the protein secretory
apparatus in bacteria, and the bacterial homologue of the eukaryotic
Signal Recognition Particle. Ffh proteins and polynucleotides are useful
CC for diagnosing diseases related to over or underexpression of Ffh protein or
CC by identifying mutations in the Ffh gene, or determining Ffh protein or
CC mRNA expression levels due to an infection of an organism with the Ffh
CC gene. They can diagnose the stage and type of infection. Ffh proteins are
CC also useful for screening for compounds which affect activity of the
protein by measuring the binding to Ffh protein and observing the
CC stimulation or inhibition of the protein function. These can be used in
CC treatment to inhibit (antagonist i.e. antibacterial drugs) or enhance
(agonist) Ffh activity, in addition to direct administration of Ffh

CC proteins to treat conditions associated with a lack of Ffh protein, or
CC direct administration of antisense sequences to prevent expression. Ffh
CC proteins (administered directly, in a vector and as a vaccine) and
CC antibodies induce an immune response to immunise and prevent disease.
CC Diseases diagnosed, prevented or treated include: bacterial infections,
CC especially Streptococcus pneumoniae infections, which cause otitis media,
CC conjunctivitis, bacteraemia, sinusitis, pleural empyema, endocarditis and
CC especially meningitis. Ffh proteins, polynucleotides and their
CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and
CC are useful for use on wounds and body implants to prevent bacterial
infection.

XX SQ Sequence 1569 BP; 479 A; 297 C; 382 G; 411 T; 0 other;

Query Match 83.3%; Score 15; DB 20; Length 1569;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gccgcgcgtcgattgct 18
|||||

Db 1196 gccgcgcgtcgattgct 1210

Search completed: April 2, 2002, 22:49:19
Job time: 9219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 08:07:11 ; Search time 165.4 Seconds
(without alignments)
24.647 Million cell updates/sec

Title: US-09-785-904-4
Perfect score: 18
Sequence: 1 catgccgtctattctg 18

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18	100.0	650	3	US-08-464-052-5
C 2	18	100.0	650	4	US-08-461-002-5
C 3	18	100.0	650	4	US-08-689-411-5
C 4	18	100.0	1535	3	US-08-464-052-1
C 5	18	100.0	1535	4	US-08-461-002-1
C 6	18	100.0	1535	4	US-08-689-411-1
C 7	18	100.0	1535	5	PCT-US94-09863-1
C 8	18	100.0	12412	1	US-08-390-878-18
C 9	18	100.0	4403765	4	US-09-103-840A-2
C 10	18	100.0	4411529	4	US-09-103-840A-1
C 11	15.4	85.6	5859	1	US-08-312-387B-1
C 12	15.4	85.6	5859	1	US-08-312-387B-7
C 13	15.4	85.6	5859	1	US-08-683-426-1
C 14	15.4	85.6	5859	1	US-08-683-426-7
C 15	15.4	85.6	5859	1	US-08-683-458-1
C 16	15.4	85.6	5859	1	US-08-683-458-7
C 17	15.4	85.6	5859	1	US-08-878-360-1
C 18	15.4	85.6	5859	2	US-08-878-360-7
C 19	15.4	85.6	5859	3	US-08-478-140B-1
C 20	15	83.3	742	1	US-08-081-072-18
C 21	15	83.3	742	1	US-08-449-093A-18
C 22	15	83.3	932	1	US-08-081-072-15
C 23	15	83.3	932	1	US-08-449-093A-15
C 24	15	83.3	1569	2	US-08-923-772-1
C 25	14.8	82.2	576	1	US-08-086-428B-18
C 26	14.8	82.2	576	1	US-08-468-570-18
C 27	14.8	82.2	576	2	US-08-290-665A-18

C 28	14.8	82.2	576	5	PCT-US95-10398-18	Sequence 18, Appl
C 29	14.8	82.2	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C 30	14.4	80.0	976	2	US-08-248-839C-5	Sequence 5, Appl
C 31	14.4	80.0	983	2	US-08-248-839C-1	Sequence 1, Appl
C 32	14.4	80.0	1200	2	US-08-248-839C-7	Sequence 7, Appl
C 33	14.4	80.0	4615	1	US-08-188-582-1	Sequence 1, Appl
C 34	14.4	80.0	4615	1	US-08-646-715-1	Sequence 1, Appl
C 35	13.8	76.7	272	1	US-08-362-670B-10	Sequence 10, Appl
C 36	13.8	76.7	272	1	US-08-362-670B-12	Sequence 12, Appl
C 37	13.8	76.7	272	3	US-08-333-576C-10	Sequence 10, Appl
C 38	13.8	76.7	272	3	US-08-333-576C-12	Sequence 12, Appl
C 39	13.8	76.7	272	4	US-08-808-324-10	Sequence 10, Appl
C 40	13.8	76.7	272	4	US-08-808-324-12	Sequence 12, Appl
C 41	13.8	76.7	272	5	PCT-US94-14030A-10	Sequence 10, Appl
C 42	13.8	76.7	272	5	PCT-US94-14030A-12	Sequence 12, Appl
C 43	13.8	76.7	926	1	US-08-362-670B-1	Sequence 1, Appl
C 44	13.8	76.7	926	3	US-08-333-576C-1	Sequence 1, Appl
C 45	13.8	76.7	926	4	US-08-808-324-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-464-052-5/c
; Sequence 5, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-5

Query Match 100.0%; Score 18; DB 3; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 catgccgtctattctg 18
Db 189 CATGCCGTCGTATTCTG 172

```
RESULT 2
US-08-461-002-5/c
; Sequence 5, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-5

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 4
US-08-464-052-1/c
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-464-052-1

Query Match 100.0%; Score 18; DB 4; Length 650;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgattgctg 18
Db 189 CATGCCGTCGTATTGCTG 172

RESULT 3
US-08-689-411-5/c
; Sequence 5, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

Query Match 100.0%; Score 18; DB 3; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgtattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 5

US-08-461-002-1/c
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

Query Match 100.0%; Score 18; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgtattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 6

US-08-689-411-1/c
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; THEREOF
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,411
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/187
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-689-411-1

Query Match 100.0%; Score 18; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 catgccgtcgtattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 7

PCT-US94-09863-1/c
; Sequence 1, Application PC/TUS9409863
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
; UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael L. Goldman
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09863
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman Mr., Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1000
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1

Query Match 100.0%; Score 18; DB 5; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18
|||||
Db 1074 CATGCCGTCGTATTGCTG 1057

RESULT 8
US-08-390-878-18/c
; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California --
; COUNTRY: USA
; ZIP: 94105

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 100.0%; Score 18; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18
|||||
Db 705 CATGCCGTCGTATTGCTG 688

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENG.H: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: (DPC 1551
; OTHER INFORMATION: "u" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18
|||||
Db 104467 catgccgtcgattgctg 104484

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 18; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 catgccgtcgattgctg 18
|||||
Db 104479 catgccgtcgattgctg 104496

RESULT 11
US-08-312-387B-1
; Sequence 1, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/683,426
;; FILING DATE:
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/312,387
;; FILING DATE: September 26, 1994
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-095B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5859 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Neisseria gonorrhoeae
;; STRAIN: F62
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..381
;; OTHER INFORMATION: glys (glycyl tRNA synthetase beta chain)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 445..1491
;; OTHER INFORMATION: lgtA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2342..3262
;; OTHER INFORMATION: lgtC
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3322..4335
;; OTHER INFORMATION: lgtD
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 4354..5196
;; OTHER INFORMATION: lgtE
;;
US-08-683-426-1

Query Match 85.6%; Score 15.4; DH 1; Length 5859;
Best Local Similarity 94.1%; Pred. No. 19;
Matches 1f; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18
||||| |||||||||

Db 2796 ATGCCGGCGTATTGCTG 2812

RESULT 14
US-08-683-426-7
; Sequence 7, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack

;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/683,426
;; FILING DATE:
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/312,387
;; FILING DATE: September 26, 1994
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-095B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5859 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Neisseria gonorrhoeae
;; STRAIN: F62
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1491..2330
;; OTHER INFORMATION: lgtB
;;
US-08-683-426-7

Query Match 85.6%; Score 15.4; DB 1; Length 5859;
Best Local Similarity 94.1%; Pred. No. 19;
Matches 1b; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgccgtcgtattgctg 18

Db 2796 ATGCCGGCGTATTGCTG 2812

RESULT 15
US-08-683-458-1
; Sequence 1, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25

Wed Apr 3 08:33:28 2002

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoeae
; STRAIN: F62
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..381
; OTHER INFORMATION: glys (glycyl tRNA syntetase beta chain)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 445..1491
; OTHER INFORMATION: lgtA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2342..3262
; OTHER INFORMATION: lgtC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3322..4335
; OTHER INFORMATION: lgtD
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4354..5196
; OTHER INFORMATION: lgtE
;
US-08-683-458-1

```

```

Query Match      85.6%; Score 15.4; DB 1; Length 5859;
Best Local Similarity 94.1%; Pred.No.19;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 atgccgcgtattgctg 18
    ||||| |||||
Db 2796 ATGCCGCGTATTGCTG 2812

```

Search completed: April 3, 2002, 08:12:53
Job time: 42533 sec

genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES

source
Location/Qualifiers

1..779
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="148A13"
/clone_lib="G"
/note="Genoscope sequence ID : COAG148AA07Lp1-end : T7"

BASE COUNT 187 a 178 c 268 g 142 t 4 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 779;

Best Local Similarity 94.4%; Pred. No. 3.6e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1;

Qy 1 catgccgtcgattgctg 18

Db 139 CATGCCGTCGTATTCCTG 122

RESULT 2

CNS02KIX 1038 bp DNA GSS 14-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 14SH15 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL201570
VERSION AL201570.1 GI:7859915
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1038)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1038)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Winkler,P., Brotier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished
AUTHORS 3 (bases 1 to 1038)
Genoscope.

TITLE Direct Submission
JOURNAL Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source
Location/Qualifiers

1..1038
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="145H15"
/clone_lib="G"
/note="Genoscope sequence ID : COAG145CD08Lp1-end : T7"

BASE COUNT 240 a 291 c 255 g 243 t 9 others
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 1038;

Best Local Similarity 94.4%; Pred. No. 3.8e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1;

Qy 1 catgccgtcgattgctg 18

Db 885 CATGCCGTCGTGTTGCTG 902

RESULT 3

AU060300/c 252 bp mRNA EST 20-MAY-1999
LOCUS AU060300 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
DEFINITION discoideum cDNA clone SLJ109, mRNA sequence.

ACCESSION AU060300

VERSION AU060300.1 GI:4881404

KEYWORDS EST.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 252)

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

TITLE Developmental cDNA in Dictyostelium discoideum

JOURNAL Unpublished (1998)

COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES Location/Qualifiers

1..252
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLJ109"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 73 a 62 c 41 g 76 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 252;

Best Local Similarity 94.1%; Pred. No. 9.2e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgtcgattgctg 18

Db 83 AAGCCGTCGTATTCG 67

RESULT 4

AU061778/c

LOCUS AU061778 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

DEFINITION discoideum cDNA clone SLF846, mRNA sequence.

ACCESSION AU061778

VERSION AU061778.1 GI:4882882

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE 1 (bases 1 to 332)

AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

TITLE Developmental cDNA in Dictyostelium discoideum

JOURNAL Unpublished (1998)

COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES

source

Location/Qualifiers

1..332
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLF846"
 /dev_stage="slug"
 94 a 92 c 49 g 97 t

BASE COUNT 94 a 92 c 49 g 97 t
 ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 332;
 Best Local Similarity 94.1%; Pred. No. 9.6e-02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcgcgtcgtattgctg 18

| | | | | | | | | | | | | | | | | | | | | |

Db 83 AAGCGTCGTATTGCTG 67

RESULT 5

N93955/c

LOCUS

DEFINITION N93955 zbf4f01.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:309337 3' similar to gb:M64497 APOLIPOPROTEIN AI REGULATORY PROTEIN-1 (HUMAN); mRNA sequence.

ACCESSION

N93955

VERSION

N93955.1

KEYWORDS

GI:1266264

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 468)

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins

M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Mardis, E., Moore

B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,

Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,

Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 480 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 223.

Location/Qualifiers

1..468

/organism="Homo sapiens"

/db_xref="GDB:1252750"

/db_xref="taxon:9606"

/clone="IMAGE:309337"

/clone_lib="Soares_fetal_lung_NbHL19W"

/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: lung; Vector: p773D (Pharmacacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGCGGCGCAATTTTTTTTTTTT-3']

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p773 vector

(Pharmacacia). Library went through one round of

normalization to a Cot ~ 5. Library constructed by Bento

FEATURES

source

Location/Qualifiers

1..332
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLF846"
 /dev_stage="slug"
 94 a 92 c 49 g 97 t

BASE COUNT 118 a 102 c 74 g 168 t
 ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 468;
 Best Local Similarity 88.9%; Pred. No. 1e+03;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 catgcgcgtcgtattgctg 18

| | | | | | | | | | | | | | | | | | | | | |

Db 383 CATNCCGTCGTATTGCTG 366

RESULT 6

AU061934/c

LOCUS

DEFINITION AU061934 Dictyostellium discoideum SL (H.Urushihiara) Dictyostellium discoideum cDNA clone SLG680, mRNA sequence.

ACCESSION

AU061934

VERSION

AU061934.1

KEYWORDS

GI:4883038

SOURCE

EST.

ORGANISM

Dictyostellium discoideum.

Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.

1 (bases 1 to 496)

REFERENCE

AUTHORS

Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,

Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,

Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

Developmental cDNA in Dictyostellium discoideum

Unpublished (1998)

Contact: Hiteko Urushihara

Institute of Biological Sciences

University of Tsukuba

3-3-10 Ten-ndai, Tsukuba, Ibaraki 305, Japan

Email: d402huesakura.cc.tsukuba.ac.jp

PROJECT = Dictyostellium discoideum cDNA project in Japan.

Location/Qualifiers

1..496

/organism="Dictyostellium discoideum"

/strain="AX4"

/db_xref="taxon:44689"

/clone="SLG680"

/clone_lib="Dictyostellium discoideum SL (H.Urushihiara)"

/dev_stage="Slug"

146 a 119 c 79 g 152 t

BASE COUNT

ORIGIN

FEATURES

source

Location/Qualifiers

1..496
 /organism="Dictyostellium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SLG680"
 /clone_lib="Dictyostellium discoideum SL (H.Urushihiara)"
 /dev_stage="Slug"

Query Match 85.6%; Score 15.4; DB 10; Length 496;
 Best Local Similarity 94.1%; Pred. No. 1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcgcgtcgtattgctg 18
 | | | | | | | | | | | | | | | | | | | | | |

Db 83 AAGCGTCGTATTGCTG 67

RESULT 7
 AU179545
 LOCUS
 DEFINITION HS_1177_A1_E04_MR Clt Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3177 Col=7 Row=I, DNA sequence.
 ACCESSION AU179545
 VERSION AU179545.1
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 508)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3177 row: I column: 7

Class: BAC ends

High quality sequence stop: 508.

Location/Qualifiers

1..508

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_plate=3177 Col=7 Row=I"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 169 a 115 c 46 g 175 t 3 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 13; Length 508;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgcgtattact 17

|||||

Db 456 CATGCCGCTGATTACT 472

RESULT 8

LOCUS BE775420 529 bp mRNA EST 20-SEP-2000

DEFINITION MY-02-G-06 pinfestansMY Phytophthora infestans cDNA, mRNA sequence.

ACCESSION BE775420

VERSION BE775420.1 GI:10229130

KEYWORDS EST.

SOURCE potato late blight agent.

ORGANISM Phytophthora infestans

Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;

Phytophthora.

1 (bases 1 to 529)

Kanoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.

Initial assessment of gene diversity for the oomycete pathogen

Phytophthora infestans based on expressed sequences

Fungal Genet. Biol. 28 (2), 94-106 (1999)

20056376

Contact: Govers F

Laboratory of Phytopathology

Wageningen University

Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands

Tel: 31 317 483 138

Fax: 31 317 483 412

Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES

source

1..529

/organism="Phytophthora infestans"

/strain="DDR7602, A1 mating type"

/db_xref="taxon:4787"

/clone_lib="pinfestansMY"

/dev_stage="4-week old vegetative, non-sporulating

mycelium in synthetic medium"

/lab_host="E. coli, strain DH5-alpha"

BASE COUNT 116 a 155 c 124 g 107 t 22 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 529;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 catgccgcgtattact 17

|||||

Db 318 CATGCCGCTGCTATTGCT 302

RESULT 9

LOCUS AW934449/c

DEFINITION EST340426 tomato fruit mature green, TAMU Lycopersicon esculentum

ACCESSION AW934449

VERSION AW934449.1 GI:8105370

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 530)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,

Upton,J., Hansen,T., Craven,M.B., Rowman,C.L., Ahn,S., Kanning,C.M.,

Fraser,C.M., Martin,C.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato fruit tissue

Unpublished (1999)

Contact: CUCI

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1..530

/organism "Lycopersicon esculentum"

/cultivar "FAL9"

/db_xref="taxon:4081"

/clone_lib="tomato fruit mature green, TAMU"

/issue_type="fruit pericarp"

/dev_stage="mature green (3-5 days pre-ripening)"

/lab_host="Solk"

/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:

XhoI; cDEF - fruit were tagged at the 1cm stage and

harvested 3-5 days prior to ripening. Fruit were cut in

half to verify the seeds were indeed 'immature' and the

seeds and locules were discarded prior to freezing the

pericarp"

BASE COUNT 164 a 117 c 103 g 146 t

ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 530;
Best Local Similarity 94.1%; Pred. NO. 1e+03;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgccgcgtattactg 18

|||||

Db 208 ATACCGCTGCTATTGCTG 192

RESULT 10

A1403760 640 bp mRNA EST 19-APR-2001
 LOCUS GH23265.5prime GH Drosophila melanogaster head pot2 Drosophila
 DEFINITION melanogaster cDNA clone GH23265 5prime, mRNA sequence.
 A1403760
 ACCESSION A1403760.1 GI:4246847
 VERSION 1
 KEYWORDS fruit fly.
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
 Lewis,S. and Rubin,G.M.
 TITLE BDGP/HMI Drosophila EST Project
 JOURNAL Contact: Stapleton, M.
 COMMENT BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd., Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 Plate: 232 row: F column: 5
 High quality sequence stop: 438.
 Location/Qualifiers
 1..640
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="GH23265"
 /sex="male and female"
 /dev_stage="adult"
 /lab_host="DH5 - alpha"
 /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
 XhoI; Sized fractionated cDNAs were directly ligated into
 pOT2; Plasmid cDNA library."
 BASE COUNT 116 a 192 c 204 g 128 t
 ORIGIN

FEATURES

source
 1..640
 2 atgcccgtcgattgctg 18
 ||||||||||||
 Db 530 AGGCGTCGTATTGCTG 546
 Query Match 85.6%; Score 15.4; DB 10; Length 640;
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 atgcccgtcgattgctg 18
 ||||||||||||
 Db 530 AGGCGTCGTATTGCTG 546

RESULT 11

BE777156 671 bp mRNA EST 20-SEP-2000
 LOCUS MY-25-B-07 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
 DEFINITION BE777156
 ACCESSION BE777156
 VERSION BE777156.1 GI:10230811
 KEYWORDS EST.
 SOURCE potato late blight agent.
 ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 Phytophthora.
 REFERENCE 1 (bases 1 to 671)
 AUTHORS Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
 TITLE Initial assessment of gene diversity for the oomycete pathogen
 Phytophthora infestans based on expressed sequences
 JOURNAL Fungal Genet. Biol. 28 (2), 94-106 (1999)
 MEDLINE 20056376
 COMMENT Contact: Govers F
 Laboratory of Phytopathology
 Wageningen University
 Binnenhaven 9, P.O.Box 8025, 6700 EH, Wageningen, The Netherlands

Tel: 31 317 483 138
 Fax: 31 317 483 412
 Email: Francine.Govers@medew.fyto.wau.nl.
 Location/Qualifiers
 1..671

FEATURES

source
 /organism="Phytophthora infestans"
 /strain="DDR7602, A1 mating type"
 /db_xref="taxon:4787"
 /clone_lib="PinfestansMY"
 /dev_stage="4-week old vegetative, non-sporulating
 mycelium in synthetic medium"
 /lab_host="E. coli, strain DH5-alpha"
 /note="Vector: pSPOT1; Site_1: SalI; Site_2: NotI; Total
 kNA was isolated from mycelium of P. infestans DDR7602
 cultured for 4 weeks in synthetic medium. EST clones were
 named by their position in the microtiter plate, preceded
 by the prefix MY (for mycelial) and the successive number
 of the microtiter plate (e.g. MY-06-A-04)."
 BASE COUNT 155 a 171 c 186 g 157 t 2 others
 ORIGIN

Query Match 85.6%; Score 15.4; DB 10; Length 671;
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcccgtcgattgctg 18

||||||||||||
 Db 429 ATTCGTCGTATTGCTG 345

RESULT 12

BE868166/c 2051 bp mRNA EST 19-JAN 2001
 LOCUS 963101A02.x1 C. reinhardtii CC-1690, Stress condition 1, normalized
 DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 BE868166
 ACCESSION BE868166.1 GI:12258310
 VERSION EST.
 KEYWORDS Chlamydomonas reinhardtii.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadales; Chlamydomonas.

Query Match 85.6%; Score 15.4; DB 10; Length 640;
 Best Local Similarity 94.1%; Pred. No. 1.1e+03;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 atgcccgtcgattgctg 18

||||||||||||
 Db 530 AGGCGTCGTATTGCTG 546

REFERENCE 1 (bases 1 to 2051)
 AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hausser,C.,
 Leclercq,E., McInerney,J.P., Shrago,J., Silliman,C. and Stern,D.
 TITLE Analysis of the Chlamydomonas Reinhardtii Genome: A Back to
 Universal System for Studying Gene Function and Regulation in
 Vascular Plants; project phase 3
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hausser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: clausen@duke.edu.
 Location/Qualifiers
 1..2051

FEATURES

source
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress condition 1,
 normalized, Lambda zap II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP S (40 min,
 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (40min, 1hr
 , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into lambda Zap II
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.

pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT

578 a 461 c 551 g 294 t 167 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 11; Length 2051;
Best Local Similarity 88.9%; Pred. No. 1.3e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 catgcgtcgtattgctg 18

|||||

Db 1664 CATGCCGTCGTCGTCG 1647

RESULT 13

LOCUS

DEFINITION RP-CI-24-333F5.TVB RPCI-24 Mus musculus genomic clone RPCI-24-333F5, DNA sequence.

ACCESSION

BH100992

VERSION

BH100992.1 GI:14925413

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

REFERENCE

AUTHORS

Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E., Russell,D., de-Jong,P. and Fraser,C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-24

JOURNAL

COMMENT

Other GSSs: RPCI-24-333F5.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from HACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 333 row: F column: 5
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..377
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="RPCI-24-333F5"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT

81 a 65 c 78 g 153 t

ORIGIN

Query Match 83.3%; Score 15; DB 13; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgccgtcgtattgct 17

|||||

Db 4 TGCCGTCGTATTGCT 18

RESULT 14

LOCUS

DEFINITION

AV834783 K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage Hordeum vulgare subsp. spontaneum cDNA clone bahl5n23, mRNA sequence.

ACCESSION

AV834783

VERSION

AV834783.1 GI:14526872

KEYWORDS

EST.

SOURCE

Hordeum vulgare subsp. spontaneum.

ORGANISM

Eukaryota; Viridiplantae; Hordeum vulgare subsp. spontaneum

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Hordeum.

1 (bases 1 to 575)

Barley EST sequencing project in NIG and Okayama Univ

Sato,K.

Unpublished (2001)

Contact: Kazuhiro Sato

Research Institute for Bioresources

Okayama University, Barley Germplasm Center

Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan

Email: kasszato@rib.okayama-u.ac.jp,

URL: <http://www.rib.okayama-u.ac.jp/barley/>

Sato,K., Saisho,D., Takeda,K., Shini,T. and Kohara,Y. Direct

submission:

database: <http://www.shigen.nig.ac.jp/barley/barley.html>.

Location/Qualifiers

1..575

/organism="Hordeum vulgare subsp. spontaneum"

/cultivar="H602"

/db.xref="taxon:77009"

/clone="bahl5n23"

/clone_lib="K. Sato unpublished cDNA library: Hordeum

vulgare subsp. spontaneum top three leaves adult, heading

stage"

/dev_stage="adult, heading stage"

/lissue_type="top three leaves"

BASE COUNT 142 a 160 c 154 g 119 t

ORIGIN

Query Match 83.4%; Score 15; DB 10; Length 575;

Best Local Similarity 100.0%; Pred. No. 1.7e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15

LOCUS

DEFINITION

AT105888 cb01a01_p1 ZF adult heart library Danio rerio cDNA 5 prime, mRNA sequence.

ACCESSION

AT105888

VERSION

AT105888.1 GI:3460991

KEYWORDS

EST.

SOURCE

zebrafish.

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Rasbora; Danio.

1 (bases 1 to 128)

Chen,J.N., deSauvage,F., Hosobuchi,M., Jackson,D.G. and Fishman

,M.C.

Expressed Sequences from The Adult zebrafish Heart

Wed Apr 3 08:33:30 2002

JOURNAL Unpublished (1998)
COMMENT Contact: Mark C. Fishman
Cardiovascular Research Center
Massachusetts General Hospital
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
Fax: 6177265806
Email: fishman@cvc.harvard.edu
http://zebrafish.mgh.harvard.edu
The original clones used for sequencing are no longer available;
the library is available from Mark C. Fishman.
Insert length: 128 Std Error: 0.00
Seq primer: p1.

FEATURES Location/Qualifiers
source 1..128
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /clone_lib="ZF adult heart library"
 /sex="mixed"
 /tissue_type="myocardium, endocardium, vessel"
 /dev_stage="adult"
 /lab_host="E. coli XL1 Blue"
 /note="Organ: heart; Vector: LambdaZAPII; Site_1: EcoRI;
 Site_2: XhoI"
BASE COUNT 23 a 36 c 29 g 31 t 9 others
ORIGIN

Query Match 82.2% Score 14.8; DB 10; Length 128;
Best Local Similarity 88.9%; Pred. No. 1.6e+03;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 catgccgtcgtattgctg-18"
 ||| |
Db 33 CAAGCGTCGTATTGCTG 50

Search completed: April 2, 2002, 21:48:05
Job time: 10830 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:49:26 ; Search time 1472.17 Seconds
(without alignments)
5076.331 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccgggtgcgcgcgcatc 453

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_om:*
- 20: em_om:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

length of match

Result No.	Score	Query Match	Length	DB	ID	Description
1	176	38.9	650	6	AR096715	AR096715 Sequence
2	176	38.9	650	6	AR147696	AR147696 Sequence
3	176	38.9	1535	6	AR096713	AR096713 Sequence
4	176	38.9	1535	6	AR147694	AR147694 Sequence
5	176	38.9	1604	1	MH035021	U35021 Mycobacteri
6	176	38.9	1453	1	MH043540	U3540 Mycobacteri
7	176	38.9	9281	1	MHDK3S1	U35017 Mycobacteri
8	176	38.9	9764	1	AE006921	AE006921 Mycobacte
9	176	38.9	10019	1	AF041819	AF041819 Mycobacte
10	176	38.9	12412	6	186264	186264 Sequence 18
11	176	38.9	17783	1	AE007028	AE007028 Mycobacte
12	176	38.9	32437	1	MTCY336	295586 Mycobacteri
13	176	38.9	38380	1	MTCY251	274410 Mycobacteri
14	176	38.9	43401	1	MTY13E12	295330 Mycobacteri
15	165	36.4	16384	1	AE007160	AE007160 Mycobacte
16	27	6.0	14179	1	AE007053	AE007053 Mycobacte
17	27	6.0	33100	1	MTCY9F9	284498 Mycobacteri
18	20	4.4	14408	1	AE005028	AE005028 Halobacte
19	20	4.4	17712	1	AE004844	AE004844 Pseudomon
20	19	4.2	9793	1	AE004944	AE004944 Pseudomon
21	19	4.2	1486	1	AE004770	AE004770 Pseudomon
22	19	4.2	28826	1	MTV041	AE021958 Mycobacte
23	19	4.2	24606	1	U70376	U70376 Streptomyce
24	19	4.2	99699	2	AC013845	AC013845 Prosophil
25	19	4.2	126340	2	DMBR18C8	AL122030 Prosophil
26	19	4.2	150621	2	AP003270	AP003270 cryza sat
27	19	4.2	151490	8	AP003611	AP003611 cryza sat
28	19	4.2	152237	8	AP001073	AE003438 Prosophil
29	19	4.2	294537	3	AE003438	AE003002 Mesorhizo
30	19	4.2	44498	1	AF003002	AF116759 Homo sapi
31	18	4.0	508	9	F116756S04	AF1245508 Colts go
32	18	4.0	521	5	CG0245508	AC057017 Giardia i
33	18	4.0	708	2	AC057017	AC077697 Giardia i
34	18	4.0	944	2	AC077697	AC030797 Giardia i
35	18	4.0	951	2	AC030797	AC083707 Giardia i
36	18	4.0	1002	2	AC083707	AF124531 Homo sapi
37	18	4.0	1193	9	HSP63G04	AX000494 Sequence
38	18	4.0	1206	6	AX000494	AX064129 Sequence
39	18	4.0	1413	6	AX064129	AX120935 Sequence
40	18	4.0	1413	6	AX120935	AX028480 Sequence
41	18	4.0	1536	6	AX028480	M12799 Corynebacte
42	18	4.0	1639	1	CG0483R	AF1360742 Brucella
43	18	4.0	1701	1	AF136074	AF114243 Corynebac
44	18	4.0	1852	1	AF114243	E00311 DNA coding
45	18	4.0	1853	6	E00311	

ALIGNMENTS

RESULT 1	AR096715	AR096715	650 bp	INA	PAT	08-SEP-2000
LOCUS	Sequence	5 from patent us 6008201.				
DEFINITION	AR096715					
ACCESSION	AR096715.1	GI:10025753				
VERSION	Unknown.					
KEYWORDS	Unclassified.					
SOURCE	1 (bases 1 to 650)					
ORGANISM	Riley, L.W.					
REFERENCE	DNA molecule encoding for cellular uptake of mycobacterium					
AUTHORS	tuberculosis and uses thereof					
TITLE	Patent: US 6008201-A 5 28-DEC-1999;					
JOURNAL	Location/Qualifiers					
FEATURES	source	1..650				
BASE COUNT	123 a	258 c	185 g	84 t		
ORIGIN						

FEATURES	source	location/Qualifiers
BASE COUNT	297 a	544 c 458 g 236 t
ORIGIN		
Query Match	38.9%;	Score 176; DB 6; Length 1535;
Best Local Similarity	100.0%;	Pred. No. 4.2e-78;
Matches	176;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	278	cgcgggcacatttgaagccgtgctagcacaactggcgccccggcgccgacaaacccg 337
DB	221	CGGGGCCACCTTTGAAGCCGTGCTAGCCAACTGGCGCGCCCGCGGACCAACCCCG 280
QY	338	acgacacaccccggtcatcgacacaccccccgatgcgccgcccgcacacgacaccc 397
DB	281	ACGACCACACCCCGGTCTATCGACACACCCCGCATGCGCGCGCCATCGACCGCACACC 340
QY	398	gcagcaagcccaacgcaacacgagcggtgctgctggcgggctgcgcgcgtgac 453
DB	341	GCAGCAAGCCCAACGCAACACGAGCGGCTGCTGCGCGGGCTGCGCGCGCTGATC 396
RESULT	2	
LOCUS	AR147696	650 bp DNA PAT 08-AUG-2001
DEFINITION	Sequence 5 from patent US 6224881.	
ACCESSION	AR147696	
VERSION	AR147696.1	GI:15111786
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	1 (bases 1 to 650)	
AUTHORS	Riley, L.W. and Chong, P.	
TITLE	DNA molecule fragments encoding for cellular uptake of	
JOURNAL	Mycobacterium tuberculosis and uses thereof	
FEATURES	Patent: US 6224881-A 5 01-MAY-2001;	
source	Location/Qualifiers	
BASE COUNT	123 a 258 c 185 g 84 t	
ORIGIN		
Query Match	38.9%;	Score 176; DB 6; Length 650;
Best Local Similarity	100.0%;	Pred. No. 4.9e-78;
Matches	176;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	278	cgcgggcacatttgaagccgtgctagcacaactggcgccccggcgccgacaaacccg 337
DB	221	CGGGGCCACCTTTGAAGCCGTGCTAGCCAACTGGCGCGCCCGCGGACCAACCCCG 280
QY	338	acgacacaccccggtcatcgacacaccccccgatgcgccgcccgcacacgacaccc 397
DB	281	ACGACCACACCCCGGTCTATCGACACACCCCGCATGCGCGCGCCATCGACCGCACACC 340
QY	398	gcagcaagcccaacgcaacacgagcggtgctgctggcgggctgcgcgcgtgac 453
DB	341	GCAGCAAGCCCAACGCAACACGAGCGGCTGCTGCGCGGGCTGCGCGCGCTGATC 396
RESULT	3	
LOCUS	AR096713	1535 bp DNA PAT 08-SEP-2000
DEFINITION	Sequence 1 from patent US 6008201.	
ACCESSION	AR096713	
VERSION	AR096713.1	GI:10025749
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	1 (bases 1 to 1535)	
AUTHORS	Riley, L.W.	
TITLE	DNA molecule encoding for cellular uptake of mycobacterium	
JOURNAL	tuberculosis and uses thereof	
Patent:	US 6008201-A 1 28-DEC-1999;	


```

CDS
/gene="MT0099.1"
3791..4162
/gene="MT0099.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MRGLRVSKVTSFAFLPTPARRAKAARVNLDQVACQRRDDEG
PTLCQCPQSARCLPTAAMSQSRVPATHCRPPCAPGAASALTCALCAEWSVVE
VRAPRAQTSFQSRAARES"
4206..4973
/gene="MT0100"
4206..4973
/gene="MT0100"
/notes="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthiodenosine
nucleosidase/S-adenosylhomocysteine nucleosidase,
putative"
/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTGVICAPOELAYLRGLVDKAKRQOVAQIILFDSQGLDAHR
VFLAAGKVNLTATLLADRGECRTIVFTGVAGLDPELCIGDIVIADRVVGHDF
GLLDERLRYQGHIPFIEPTERLGYVPDPAVIDRVKHLRDLGFTLAPLSTAAGGGGR
QPIYYITGLTGOOYLHCERTRNRLHIEIGMAVEMEGGNAVQICASFDPILWLIVRAL
SDLAGDSGVDFNRFYGEVAASSARVLLRLPLVLTAC"
5105..7390
/gene="MT0101"
5105..7390
/gene="MT0101"
/notes="similar to SP:P46840 PTD:559912; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTTAVTGEHHIASVORIQLRISGMSCSACARHVESTLNKLPGVRA
AVNFGTRVATIDTSEAVDAALQAVRRAGTQADLDDGRSASDPDADHARQLLILK
AIAAVLVGVADVPVFPATFTGWMVLSALALPVVTWMAAPHRVAMRNARHH
AASMETILSVGITAATTWISLYTFVGNHSP IERSG IQOALLGSDA [YF:EVAAGVTVPVL
VGRYFEARAKSOAALRAALAAI.SAKEVAVILPDGSEMWI PAIDILKQQRFFVVRPQI
VAADGLAVDGSAAVDMSAMTGEAKPTRVRPGGVIGCTTVLLDGLI.IVEAAVACADTQF
AGNRLVEQQAQADARLADRLSSVFVPVAVLVAALTAAAGMLIAGQPDRAVSAAI
AVLVICPCALGLATPTAMVAVSGRGAQIGIFLKGYSLEATRAVUTVVFDFKTTLTT
GRLOVSATVAPGWEADOVIALAATVEAASEHVALAIAAATTHRDVATDFRALPGRG
VSGTVSGRAVRGKPSWIGSSCHPNMRAARRHAESLGETAVPFVVDGEPGCVIADAV
AVKDSADVAALADRGRLTMLITGDNPESAVAATRVGIDEVIADILPEGKVDVIEQ
LDRGHVVMVGDGINDGPALARADLQMAIGRTDVAIGAADIILVRDHLDDVVPALAD
LARATMTVKLNMVAFGYNTAATPVAAAGLLNPLVAGAAAFSFFVWNSLRKRF
GRYPLGCGTGGPQMTAPSSA"
complement(7337..8185)
/gene="MT0102"
complement(7337..8185)
/gene="MT0102"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44325.1"
/db_xref="GI:13879149"
/translation="MLAQATTAGSFNHHASTVLOGCRGVPAAMWSEPAQATRRHCATI
DGMDCVEARLSALDGERAPVPSRDEHLGECRSACRAWTFQVAGSDGLRRLAES
RPVPPVGRGLIRRAPRROHSPMTWRWALLCVIAQIALCTVCGFLDVLTHOHP
GAGTHLLNESTSWSLALCVIMVGAALMPSAAAGLAGVLTAFVALTCYVILVDALSGAV
STTRILTLPPVIVGAVLAIMVWRASGPRPRPDVAAPEDVILPDNARSRRGRHLNP
TDGSAA"
complement(8232..9095)
/gene="MT0103"

gene
/gene="MT0099.1"
3791..4162
/gene="MT0099.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879150"
/translation="NMULHPDGLDUTERARKRGITLSNOOYDGMRSKLSGYLTPOARA
TFEAVLAKLAAPCATNDDHPVIDTTPDAAALDRDTSOAKRNDHGLIAGLRALIAS
KICQGHNLPLPSIVVTTLTLDQTCAGKGTGGGTLLPMADYIKMTSHAHYSPASGR
YPAJIFDGHGTPLALYITRLASPAPQIRMLFANDRGCTRGCDAPAYHISQAHIVTAWTS
TURTDITELTLACGPDNRLAERKGTTHKNTHGHTMLPPHILDHGOI:HTNTEFHHPERF
LHNQDDDDKPD"
complement(8956..9623)
/gene="MT0104"
/notes="this region contains an authentic frame shift and
is not the result of a sequencing artifact; identified by
Glimmer2; putative; conserved hypothetical protein,
authentic frameshift"
BASE COUNT 1553 a 3103 c 3360 g 1748 t
ORIGIN

Query Match 38.9%; Score 176; DB 1; Length 9764;
Best Local Similarity 100.0%; Pred. No. 2.9e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 CQCGGCGCACCTTGAAGCCYCTACTAGCAAACTGCGCCCGCGCGGCGACCAACCCY 437
|||||
Db 8971 CGCGGCGCACCTTGAAGCCGCTGCTAGCAAACTGCGCCCGCGCGGCGACCAACCCG 8912
|||||

QY 338 acgactacacccgggtcatcgacactaccccgatgugcgcccatcgacgcgacaccc 397
|||||
Db 8911 ACGACCAACACCCCGTTCATCCACACACCCCGATGCGCGCGCATCCACCGGACACCC 8852
|||||

QY 398 gcagccaagcccaacgcaacacgacagcgctgctggtgcccgggtgcgcgcctatc 453
|||||
Db 8851 GCAGCCAAGCCCAACGCAACACGACGCGGTGCTGCGCGCGGTGCGCGCGCTGATC 8796
|||||

RESULT 9
AF041819/c AF041819 10019 bp DNA BCT 18-JAN-1998
LOCUS Mycobacterium bovis BCG; putative
DEFINITION adenosylmethionine-7-oxononanoate aminotransferase (bioA),
putative 8 amino-7-oxononanoate synthase (bioF), putative
dethiobiotin synthetase (bioB), and biotin synthetase (bioB) genes,
complete cds.
ACCESSION AF041819
VERSION AF041819.1 GI:2791842
KEYWORDS Mycobacterium bovis BCG.
SOURCE Mycobacterium bovis BCG.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 10019)
AUTHORS Yu, S. and Jacobs, W.R. Jr.
TITLE Cloning, sequencing, and identification of Mycobacterium bovis BCG
biotin biosynthetic genes by complementing two Mycobacterium
smegmatis biotin mutants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10019)
AUTHORS Yu, S.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Microbiology & Immunology, Albert Einstein
College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10010, USA
FEATURES
source
1..10019
/organism="Mycobacterium bovis BCG"
/strain="Pasteur"
/sub_species="BCG"
```

CDS

```

complement(8232..9095)
/gene="MT0103"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44326.1"
/db_xref="GI:13879150"
/translation="NMULHPDGLDUTERARKRGITLSNOOYDGMRSKLSGYLTPOARA
TFEAVLAKLAAPCATNDDHPVIDTTPDAAALDRDTSOAKRNDHGLIAGLRALIAS
KICQGHNLPLPSIVVTTLTLDQTCAGKGTGGGTLLPMADYIKMTSHAHYSPASGR
YPAJIFDGHGTPLALYITRLASPAPQIRMLFANDRGCTRGCDAPAYHISQAHIVTAWTS
TURTDITELTLACGPDNRLAERKGTTHKNTHGHTMLPPHILDHGOI:HTNTEFHHPERF
LHNQDDDDKPD"
complement(8956..9623)
/gene="MT0104"
/notes="this region contains an authentic frame shift and
is not the result of a sequencing artifact; identified by
Glimmer2; putative; conserved hypothetical protein,
authentic frameshift"
BASE COUNT 1553 a 3103 c 3360 g 1748 t
ORIGIN

Query Match 38.9%; Score 176; DB 1; Length 9764;
Best Local Similarity 100.0%; Pred. No. 2.9e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 CQCGGCGCACCTTGAAGCCYCTACTAGCAAACTGCGCCCGCGCGGCGACCAACCCY 437
|||||
Db 8971 CGCGGCGCACCTTGAAGCCGCTGCTAGCAAACTGCGCCCGCGCGGCGACCAACCCG 8912
|||||

QY 338 acgactacacccgggtcatcgacactaccccgatgugcgcccatcgacgcgacaccc 397
|||||
Db 8911 ACGACCAACACCCCGTTCATCCACACACCCCGATGCGCGCGCATCCACCGGACACCC 8852
|||||

QY 398 gcagccaagcccaacgcaacacgacagcgctgctggtgcccgggtgcgcgcctatc 453
|||||
Db 8851 GCAGCCAAGCCCAACGCAACACGACGCGGTGCTGCGCGCGGTGCGCGCGCTGATC 8796
|||||

RESULT 9
AF041819/c AF041819 10019 bp DNA BCT 18-JAN-1998
LOCUS Mycobacterium bovis BCG; putative
DEFINITION adenosylmethionine-7-oxononanoate aminotransferase (bioA),
putative 8 amino-7-oxononanoate synthase (bioF), putative
dethiobiotin synthetase (bioB), and biotin synthetase (bioB) genes,
complete cds.
ACCESSION AF041819
VERSION AF041819.1 GI:2791842
KEYWORDS Mycobacterium bovis BCG.
SOURCE Mycobacterium bovis BCG.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 10019)
AUTHORS Yu, S. and Jacobs, W.R. Jr.
TITLE Cloning, sequencing, and identification of Mycobacterium bovis BCG
biotin biosynthetic genes by complementing two Mycobacterium
smegmatis biotin mutants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10019)
AUTHORS Yu, S.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Microbiology & Immunology, Albert Einstein
College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10010, USA
FEATURES
source
1..10019
/organism="Mycobacterium bovis BCG"
/strain="Pasteur"
/sub_species="BCG"
```

CDS	<p>/db_xref="taxon:33892" complement(259..1038) /note="ORF1; similar to MTCY336.37" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAB96954.1" /db_xref="GI:2791843" /translation="MKRSMKSGFAIGLAMLPVMAAPGLAAADPATRPVDYQQITD VYIARGLSORGVPSWAGGIGSPTRGTGTGINTVDFDASGLIYAYAGAGLKLPRSS COMTKYGGKVLPOQARKGDLIFVCGPTQSVALLYLKGOMLEVGDDVVOVSPVRTNGMT PYLVFGSPSRPNRRKRSQRRSKRPSKSGARPTGAGTGA GPASAPASAR PAASLRHALTITVPLAFSPKPAFCPLLSARSRLVWN" complement(1278..1850) /note="ORF2; similar to MTCY336.36" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAB96955.1" /db_xref="GI:2791844" /translation="MICSQVNPVVAAMEIEGNAATPRPIPVTVAKTSAIGEHACMAP PNOPEDTQDAGTESRSALDEPDBERTPPRVLTIRIAPWDVVCYVILAVFLVMYMTS WPSRLFAFTDNVCPDACPVPFGVNYIYIPVMNGGICAAIATATVIGPFVSMKGYTM SEWPIISTAVITVTSIAGYATAGFSERYWH" 1810..3123 /gene="bioA" 1810..3123 /gene="bioA" /note="BioA; similar to MTCY336.35c" /codon_start=1 /transl_table=11 /product="putative adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AAB96956.1" /db_xref="GI:2791845" /translation="NAAATGGTLTPEQIIAIVDGAHLWHPYSSIGKEAVSPVVAAIIGA WILTRDGOPIELVDMSSWNTA IHGHGHPALDQALTOLKVMNHVMEGGLTHEPPAAR LAKLIVITPAGLDIVFSDSGSVSVEAAKALQYWRGRGLPCKRKLMTWRGQVHGD TELAMISICPHGMHSLMTDVLAAQVFAPOVPRDYDPAYSAFAQLAQHAGELAAVV VEPVQAGMHPDPRYLHDLRDLCRRIVELLIIIDEIATGFGRTGALFAADHAGVSP DIMCVKALTGGYLSLAATLTADYVAITISAGAAGALMHGPTFMANPLACAVSVASVE LLLGQDMRTITELAAGLTAGLDTARALPAPVDVRCVGAIGVIECDRPVDLAVATPAA LDRGVYLRPRFNLVAMPPICTPAEITQITSAMVEVARLVGSLP" 3120..4280 /gene="bioF" 3120..4280 /gene="bioF" /EC_number="2.3.1.47" /note="BioF; similar to MTCY336.34c" /codon_start=1 /transl_table=11 /product="putative 8-amino-7-oxononanoate synthase" /protein_id="AAB96957.1" /db_xref="GI:2791846" /translation="MKAATQARIDDSPLAWLDVAVQORHEAGIRKCLRP RPRAVATELD LAGNDYLGLSRHPAVIDGGVQALRIWAGATGSRIVTGDKLHQOFEAPLEAFVGAAA GLFSSGYTANIGAVVGLSGPGLLVSDARSHASLVDACRLSRARVVVTPHRDVAVD AALSRDEOPRVVTDVSFVSADGSLAPVRELLEVCRRHGALLVDEAHGIGVVRGGGRG LLYELGLAGAPDVVWTTLSKALGSGGVVLTGPVVRHLIDAARPFIFDTGLAPAAY GAARAALRQAEPWP RPQAVLNVHAGELARMCGVAAPVDSAMYSVILGEPSAVAAAA CLDAGYKVGCGFRPPTVPAGTSRLRLTARASLNAGLELELARRVLTDLAVARR" 4277..4957 /gene="bioD" 4277..4957 /gene="bioD" /EC_number="6.3.3.3" /note="BioD; similar to MTCY336.33c" /codon_start=1 /transl_table=11 /product="putative dethiobiotin synthetase" /protein_id="AAB96958.1" /db_xref="GI:2791847"</p>	
CDS	<p>/db_xref="taxon:33892" complement(259..1038) /note="ORF1; similar to MTCY336.37" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAB96954.1" /db_xref="GI:2791843" /translation="MKRSMKSGFAIGLAMLPVMAAPGLAAADPATRPVDYQQITD VYIARGLSORGVPSWAGGIGSPTRGTGTGINTVDFDASGLIYAYAGAGLKLPRSS COMTKYGGKVLPOQARKGDLIFVCGPTQSVALLYLKGOMLEVGDDVVOVSPVRTNGMT PYLVFGSPSRPNRRKRSQRRSKRPSKSGARPTGAGTGA GPASAPASAR PAASLRHALTITVPLAFSPKPAFCPLLSARSRLVWN" complement(1278..1850) /note="ORF2; similar to MTCY336.36" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAB96955.1" /db_xref="GI:2791844" /translation="MICSQVNPVVAAMEIEGNAATPRPIPVTVAKTSAIGEHACMAP PNOPEDTQDAGTESRSALDEPDBERTPPRVLTIRIAPWDVVCYVILAVFLVMYMTS WPSRLFAFTDNVCPDACPVPFGVNYIYIPVMNGGICAAIATATVIGPFVSMKGYTM SEWPIISTAVITVTSIAGYATAGFSERYWH" 1810..3123 /gene="bioA" 1810..3123 /gene="bioA" /note="BioA; similar to MTCY336.35c" /codon_start=1 /transl_table=11 /product="putative adenosylmethionine-8-amino-7-oxononanoate aminotransferase" /protein_id="AAB96956.1" /db_xref="GI:2791845" /translation="NAAATGGTLTPEQIIAIVDGAHLWHPYSSIGKEAVSPVVAAIIGA WILTRDGOPIELVDMSSWNTA IHGHGHPALDQALTOLKVMNHVMEGGLTHEPPAAR LAKLIVITPAGLDIVFSDSGSVSVEAAKALQYWRGRGLPCKRKLMTWRGQVHGD TELAMISICPHGMHSLMTDVLAAQVFAPOVPRDYDPAYSAFAQLAQHAGELAAVV VEPVQAGMHPDPRYLHDLRDLCRRIVELLIIIDEIATGFGRTGALFAADHAGVSP DIMCVKALTGGYLSLAATLTADYVAITISAGAAGALMHGPTFMANPLACAVSVASVE LLLGQDMRTITELAAGLTAGLDTARALPAPVDVRCVGAIGVIECDRPVDLAVATPAA LDRGVYLRPRFNLVAMPPICTPAEITQITSAMVEVARLVGSLP" 3120..4280 /gene="bioF" 3120..4280 /gene="bioF" /EC_number="2.3.1.47" /note="BioF; similar to MTCY336.34c" /codon_start=1 /transl_table=11 /product="putative 8-amino-7-oxononanoate synthase" /protein_id="AAB96957.1" /db_xref="GI:2791846" /translation="MKAATQARIDDSPLAWLDVAVQORHEAGIRKCLRP RPRAVATELD LAGNDYLGLSRHPAVIDGGVQALRIWAGATGSRIVTGDKLHQOFEAPLEAFVGAAA GLFSSGYTANIGAVVGLSGPGLLVSDARSHASLVDACRLSRARVVVTPHRDVAVD AALSRDEOPRVVTDVSFVSADGSLAPVRELLEVCRRHGALLVDEAHGIGVVRGGGRG LLYELGLAGAPDVVWTTLSKALGSGGVVLTGPVVRHLIDAARPFIFDTGLAPAAY GAARAALRQAEPWP RPQAVLNVHAGELARMCGVAAPVDSAMYSVILGEPSAVAAAA CLDAGYKVGCGFRPPTVPAGTSRLRLTARASLNAGLELELARRVLTDLAVARR" 4277..4957 /gene="bioD" 4277..4957 /gene="bioD" /EC_number="6.3.3.3" /note="BioD; similar to MTCY336.33c" /codon_start=1 /transl_table=11 /product="putative dethiobiotin synthetase" /protein_id="AAB96958.1" /db_xref="GI:2791847"</p>	
gene	<p>7455..8504 /gene="bioH" 7455..8504 /gene="bioH; similar to MTCY336.15c" /codon_start=1 /transl_table=11 /product="biotin synthetase" /protein_id="AAB96962.1" /db_xref="GI:2791851" /translation="MTQAATPRPTNADGQGNNSDILLVARQOVLOKQKGINODOVILA VIQLPDDRLELLALAHFVRMRWCGPEVEGIIISLKTGGCPELCHPVSOSGLFPASPV RSAWLDIPSLVEAAKQAKSGATEFCIVAAGVGPDERLMAQOVAAGIETAIKNEVEINIA CSICMLTAARQVDOLAAARGVIIRYNNHNLETSFFANVYTHITWEERWGTIISMVRDAGME VCCGMLTGMGTLOOARAEFAELAEGLGDPVPLNLPNRPCTPFADILEVMPVGDALKA VAARIALPRTMIKFAAGGRITLGDIGAKKGILGGINAVIVGNVLTTLTGHPAEADLEI. IDELQMLPKALNASI." 8825..9406 /note="ORF10; similar to MTCY336.13c" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAB96963.1" /db_xref="GI:2791852" /translation="MLGSATGVLVGGWLWMIAPPIHVAIVAITRAGERVHEYLGSSESQ NFFIAPFMLLGLLSVLYAVASALMQWREHGPQWVAGLSIGLITAAAIAGVAGLVV RLRYGALDFDTVPILSRGDHALTYVTQAPVFFARRPQIALTLMMPAGIASLVVALLA ACTARDDLGGYPVDPSPSSNARTELETQAPVS" complement(9571..>10019) /note="ORF11; similar to MTCY336.12" /codon_start=3 /transl_table=11</p>	

```
/product="unknown"
/protein_id="AAB96964.1"
/db_xref="GI:2791853"
/translation="RMAGROMDFLEDEPILSTPEVSHVFGDTKLGSAVPPPPVLI
VQAVHDLIVSDIDALADSYTAGGANVTHRDLFSEHSLHPLSAPMTLRLWLDRA
GKPLTDHRVTRFTWPTIFNPMYTAGMARLAVIAAKVITGRKLSRPL"
BASE COUNT      1456 a  3134 c  3655 g  1774 t
ORIGIN

Query Match      38.9%; Score 176; DB 1; Length 10019;
Best Local Similarity 100.0%; Pred. No. 2.9e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcggccaccttgaagccgtgtagcacaactgcccgcggccgcgcgacaccccg 337
|||||
DB 6354 CGCGGGCCACCTTTGAAGCCGTGCTAGCAAACTGCGCGCGCGCGCGACCAACCCCG 6295
|||||

QY 338 acgacacaccccggtatcgacacacaccccgatcgcgccgcacatcgacgcgacaccc 397
|||||
DB 6294 ACGACCACACCCCGTATCGACACACCCCGGATGCGCGCGCATCGACCGCGACACCC 6235
|||||

QY 398 cgagcaagcccaacgcaacacgacgagcggtgctgagcgggctgcgcgcgtgac 453
|||||
DB 6234 GCAGGCAAGCCCAACGCAACACGACGAGCGCTGCTGCGCGGCGTGCAGCGCTGATC 6179
|||||

RESULT 10
186264
LOCUS      186264      12412 bp      DNA      PAT      10-JUN-1998
DEFINITION Sequence 18 from patent US 5700683.
ACCESSION 186264
VERSION   186264.1 GI:3205982
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 12412)
AUTHORS   Stover,C,Kendall and Mahairas,G.G.
TITLE     Virulence-attenuating genetic deletions deleted from mycobacterium
BCG
JOURNAL   Patent: US 5700683-A 18 23-DEC-1997;
FEATURES   Location/Qualifiers
            source      1..12412
                /organism="unknown"
BASE COUNT 2146 a  4274 c  3875 g  1917 t  200 others
ORIGIN

Query Match      38.9%; Score 176; DB 6; Length 12412;
Best Local Similarity 100.0%; Pred. No. 2.7e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcggggcaccttgaagccgtgtagcacaactgcccgcggccgcgcgacaccccg 337
|||||
DB 737 CGCGGGCCACCTTTGAAGCCGTGCTAGCAAACTGCGCGCGCGCGCGACCAACCCCG 796
|||||

QY 338 acgacacaccccggtatcgacacacaccccgatcgcgccgcacatcgacgcgacaccc 397
|||||
DB 797 ACGACCACACCCCGTATCGACACACCCCGGATGCGCGCGCATCGACCGCGACACCC 856
|||||

QY 398 cgagcaagcccaacgcaacacgacgagcggtgctgagcgggctgcgcgcgtgac 453
|||||
DB 857 GCAGGCAAGCCCAACGCAACACGACGAGCGCTGCTGCGCGGCGTGCAGCGCTGATC 912
|||||

RESULT 11
AE007028/c
LOCUS      AE007028      17783 bp      DNA      BCT      27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 114 of the
complete genome.
ACCESSION  AE007028 AE000516
VERSION    AE007028.1 GI:13881250
```

```
KEYWORDS      Mycobacterium tuberculosis CDC1551.
SOURCE        Mycobacterium tuberculosis CDC1551
ORGANISM      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Acidimicrobiales; Corynebacterineae; Mycobacteriaceae;
              Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE     1 (bases 1 to 17783)
AUTHORS      Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
              Peterson,J., DeBoy,R., Dodson,R., Winn,M., Haft,D., Hickey,E.,
              Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
              Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
              Gill,J., Mikula,A. and Bishai,W.
              Whole genome comparison of Mycobacterium tuberculosis clinical and
              laboratory strains
              Unpublished
              2 (bases 1 to 17783)
              Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
              Peterson,J., DeBoy,R., Dodson,R., Winn,M., Haft,D., Hickey,E.,
              Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
              Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
              Gill,J., Mikula,A. and Bishai,W.
              Direct Submission
              Submitted (25-APR-2001) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA
              Location/Qualifiers
                source      1..17783
                    /organism="Mycobacterium tuberculosis CDC1551"
                    /strain="CDC1551"
                    /db_xref="taxon:83331"
                    /note="clinical strain"
                    /complement(104..826)
                    /gene="MT1622.1"
                    /complement(104..826)
                    /gene="MT1622.1"
                    /note="identified by Glimmer2; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAK45890.1"
                    /db_xref="GI:13881251"
                    /translation="MLAKLAAPGATNPDDHTPVIDTPDAAAIIDRTKSOQRNHDGL
                    IAGIKALIASGKIAOHNGLPVSIVVTTLTDLQTCAGKPGTGGTLLIPMAIVIRMTSH
                    AHHYSPASGRYPQAIIDHGTPLALYHTKRIASPAORIMLFANDKGCYKPGCIAPAYHS
                    QAHVTAWTSGRTFTDTHACDPINRLAKGWTHTKNTHTGHTIEMLPPIHLDHCQPR
                    INPFIHHEKLLKHNPENHNDP"
                    /complement(841..1499)
                    /gene="MT1624"
                    /complement(841..1499)
                    /gene="MT1624"
                    /note="identified by Glimmer2; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAK45891.1"
                    /db_xref="GI:13881252"
                    /translation="MLANSHEELVEVFDALDAELDRDEVSFEVLTTPERIRSLERLE
                    CLVRLPVAGHTLNLQDQTQASEELGTLCCALANRLRITKPDAAIRIADAADLGR
                    RAITGEPLAQLTATATAGQGLIGEAHIVRALFRPPARRGVCVPIPPGRRSPGRQ
                    SRSTSSRRAGPLRPAGHGLATPRRRPHRHTRPQTRHHPEQPAIRKIIVTAKWLDPDPS
                    AGH"
                    1947..2996
                    /gene="MT1624"
                    1947..2996
                    /gene="MT1624"
                    /note="similar to GP:1680650; identified by sequence
                    similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="biotin synthase"
                    /protein_id="AAK45892.1"
                    /db_xref="GI:13881253"
                    /translation="MTQAATPPTNDAGDCGNSDILVVARQGVLOHCEGILNDDVLA
                    VLQIPDDRLELLALAEHVMRWCCGPEVEVEGIIISLKTGGCPPEICHTPSQSLASPV
```


KEYWORDS
SOURCE
ORGANISM

Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS

1 (bases 1 to 32437)
Cole, S.T., Brosch, R., Parkhill, J., Garner, T., Churcher, C.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellern, S., and
Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)

TITLE

2 (bases 1 to 32437)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75224 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2117233.

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES
source

Location/Qualifiers
1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..273

source

/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y48"
33..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y336"

source

complement(39..2228)
/gene="Rv1565c"
complement(39..2228)
/gene="Rv1565c"
/note="Rv1565c"
protein, some similarity to O05402 HYPOTHETICAL 72.2 KD
PROTEIN from B. subtilis (634 aa) opt: 384 E(): 4.8e-17;
29.1% identity in 378 aa overlap and to hypothetical
protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, Y392_HAFIN P43993 h10392 (245 aa) opt: 265;
E(): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse
score is 0.930"

gene

CDS

complement(2327..3019)
/gene="Rv1566c"
complement(2327..3019)
/gene="Rv1566c"
/note="Rv1566c" (MTCY336.37), len: 230. Function: unknown
probably exported has QOAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasins
1,2(O07390, O07391) from M. avium and slightly similar to
C-terminus of *Listeria* invasion-associated protein p60
precursor p60.LISMO P21171. FASTA results, Q49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. TBparse score is 0.890"

/transl_table=11
/product="hypothetical protein Rv1565c"
/protein_id="CAB09072.1"
/db_xref="GI:2117272"
/db_xref="SPTREMBL:O06625"
/translation="MLTSPRRPALTPPEALPVTMGTRTTGTFYRHLDGURGVAlA
LVAVPHWFGVSGGVDFLALSGFFGKILRAALNPDLSPISPAEIVRLIRLLPA
LVVLACALLTIAIOPTRWEAFANOSLASGYQNNELASTVSNYLRAGAVSPLO
HWMSVQCOFYLAFLLLVAGGACLLRLFRGPRAPYLRTMFVLLSTITIASFYAI
VHAHQAQATYNTFAKELMELAGALVGVPHVPHWMLRTAVATAALATLSCGAL
IDGKFFPGPMALVPVGCATMLAGANQGHGPTGDRPLPLNRLVEDLRLRAPGV
YSYLWHMPLLIPLSYTHRHANFVTEAAILVGLLALYLRLLVEDLRLRAPGV
RSPAAVPPIPWRLRRPTVLGSAVALTATSTFWREHVIQVRAAGKELSGLS
SRDPCARALIDHVRPKLMRPTVLELZDLPTSTKDCISDFVNPALINCYGSDV
APTIALAGSHAHEHLTALDLGRMHFKVVTYLMKGCPLSTEEVPLIMGNAPYPO
CHOWQAAMAKLVADHPDYVFTSTRPNKIPGDVMPATYVGIWOTFADNNTPLAMR
DTWLVKDGOPFIPADCLAKGNQSCGIARSKVLVDRNPTLDFAVKPFLKPLDMSD
AICRTDTCRAVEGNVLVIRDSHLLTPTTMRWTSTSELGKQIAANTDMW"
complement(2327..3019)
/gene="Rv1566c"
complement(2327..3019)
/gene="Rv1566c"
/note="Rv1566c" (MTCY336.37), len: 230. Function: unknown
probably exported has QOAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasins
1,2(O07390, O07391) from M. avium and slightly similar to
C-terminus of *Listeria* invasion-associated protein p60
precursor p60.LISMO P21171. FASTA results, Q49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. TBparse score is 0.890"

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAH09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:O06624"
/translation="MKRSMKSGSFATGLAMVLAPMVAAPGLAAADPATRPVDVQOITD
VVIARLQSGVPSWAGGIGISGTGTGTGINTVGFDSGLLIQYAYAGAGLIPRSS
COMYKVQKVLPOQARKDLIFYPGEGTSVALYLGKQMLEVDGVVQVSPVPTNGMT
PYLVRLGTQPTPVQAPVPAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAP
VOPPFCTARS"
complement(3030..3036)
/note="possible RBS. AGGAGG, for Rv1566c"
complement(3258..3263)
/note="ASNI site: ATTAAT; probably linking fragments
H3/G6"
complement(3259..3543)
/note="Rv1566c"
complement(3259..3543)
/gene="Rv1567c"
/note="Rv1567c" (MTCY336.36), len: 94. Unknown. TBparse
score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:O06623"
/translation="MVTMTSPSLPFAFTDNVCPDAPLVPFGVNYLYPVMWGGIG
AAIATAVIGPVPVSMKLGWYMFNPIISIAVITVTSIAGYATAGFSERLWH"
3791..5104
/gene="bioA"
3791..5104
/gene="bioA"
/note="Rv1568" (MTCY336.35c), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC:6.1.62) O06622. Contains
aminotransferase class-III pyridoxal-phosphate attachment
site (P500600). FASTA results, BIOA_MYCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
identity in 421 aa overlap. TBparse score is 0.874"

[illegible]

VERSION	2741410.1	GI:3261600																																										
KEYWORDS	Mycobacterium tuberculosis H37Rv.																																											
SOURCE	Mycobacterium tuberculosis H37Rv																																											
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.																																											
REFERENCE:	1 (bases 1 to 38380)																																											
AUTHORS	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, K., Devlin, K., Feldwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrall, B.G.																																											
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence																																											
JOURNAL	Nature. 393 (6685), 537-544 (1998)																																											
MEDLINE	98295987																																											
REFERENCE	2 (bases 1 to 38380)																																											
AUTHORS	Parkhill, J.																																											
TITLE	Direct Submission																																											
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk																																											
COMMENT	On Jun 27, 1998 this sequence version replaced gi:1405752. Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of 78 genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.																																											
FEATURES	<table border="0"> <tr> <td>location/qualifiers</td><td></td></tr> <tr> <td>1..38380</td><td></td></tr> <tr> <td>/organism="Mycobacterium tuberculosis H37Rv"</td><td></td></tr> <tr> <td>/strain="H37Rv"</td><td></td></tr> <tr> <td>/db_xref="taxon:83342"</td><td></td></tr> <tr> <td><1..35889</td><td></td></tr> <tr> <td>/organism="Mycobacterium tuberculosis H37Rv"</td><td></td></tr> <tr> <td>/strain="H37Rv"</td><td></td></tr> <tr> <td>/db_xref="taxon:83332"</td><td></td></tr> <tr> <td>/clone="Y251"</td><td></td></tr> <tr> <td>6..956</td><td></td></tr> <tr> <td>/gene="hycD"</td><td></td></tr> <tr> <td>6..956</td><td></td></tr> <tr> <td>/gene="hycD"</td><td></td></tr> <tr> <td>/note="Rv0084, (MTCY251.02), len: 316, hycD, similar to HCD_ECOLI P16430 formate hydrogenlyase subunit 4 (307 aa) FASTA scores, opt: 705, z-score: 591.6, E(): 2.1e-26, (33.8% identity in 305 aa overlap) and NUOH_ECOLI p33603nadh dehydrogenase i chain h (325 aa) FASTA scores, opt: 207, z-score: 220.7, E(): 9.5e-06, (26.5% identity in 260 aa overlap)"</td><td></td></tr> <tr> <td>/codon_start=1</td><td></td></tr> <tr> <td>/transl_table=11</td><td></td></tr> <tr> <td>/product="hycD"</td><td></td></tr> <tr> <td>/protein_id="CAA98920.1"</td><td></td></tr> <tr> <td>/db_xref="GI:1405754"</td><td></td></tr> <tr> <td>/db_xref="SPTREMBL:O10881"</td><td></td></tr> </table>		location/qualifiers		1..38380		/organism="Mycobacterium tuberculosis H37Rv"		/strain="H37Rv"		/db_xref="taxon:83342"		<1..35889		/organism="Mycobacterium tuberculosis H37Rv"		/strain="H37Rv"		/db_xref="taxon:83332"		/clone="Y251"		6..956		/gene="hycD"		6..956		/gene="hycD"		/note="Rv0084, (MTCY251.02), len: 316, hycD, similar to HCD_ECOLI P16430 formate hydrogenlyase subunit 4 (307 aa) FASTA scores, opt: 705, z-score: 591.6, E(): 2.1e-26, (33.8% identity in 305 aa overlap) and NUOH_ECOLI p33603nadh dehydrogenase i chain h (325 aa) FASTA scores, opt: 207, z-score: 220.7, E(): 9.5e-06, (26.5% identity in 260 aa overlap)"		/codon_start=1		/transl_table=11		/product="hycD"		/protein_id="CAA98920.1"		/db_xref="GI:1405754"		/db_xref="SPTREMBL:O10881"	
location/qualifiers																																												
1..38380																																												
/organism="Mycobacterium tuberculosis H37Rv"																																												
/strain="H37Rv"																																												
/db_xref="taxon:83342"																																												
<1..35889																																												
/organism="Mycobacterium tuberculosis H37Rv"																																												
/strain="H37Rv"																																												
/db_xref="taxon:83332"																																												
/clone="Y251"																																												
6..956																																												
/gene="hycD"																																												
6..956																																												
/gene="hycD"																																												
/note="Rv0084, (MTCY251.02), len: 316, hycD, similar to HCD_ECOLI P16430 formate hydrogenlyase subunit 4 (307 aa) FASTA scores, opt: 705, z-score: 591.6, E(): 2.1e-26, (33.8% identity in 305 aa overlap) and NUOH_ECOLI p33603nadh dehydrogenase i chain h (325 aa) FASTA scores, opt: 207, z-score: 220.7, E(): 9.5e-06, (26.5% identity in 260 aa overlap)"																																												
/codon_start=1																																												
/transl_table=11																																												
/product="hycD"																																												
/protein_id="CAA98920.1"																																												
/db_xref="GI:1405754"																																												
/db_xref="SPTREMBL:O10881"																																												
SOURCE	<table border="0"> <tr> <td>source</td><td></td></tr> <tr> <td>gene</td><td></td></tr> <tr> <td>CDS</td><td></td></tr> </table>		source		gene		CDS																																					
source																																												
gene																																												
CDS																																												

```
/translation="MSYLAGAAQICGVWVCAPIVIGMTROVRAHWEGRAGAGLLQPHR
DLKQLGQQOITPAGTITIVFAAAPVIVAGITITLLLAPELPLVATGSPDPADPLFAYVG
LIFTGALTLAGIDTGTSGFGMGASREITIAALVEPTILLAVFALSIPAGSANLGAAL
VASTIDHCHVYSLAGVLAVFALVIVIAETGRPLVPDNPATHLTLTWHEAMVLEYAG
PRLAWEAGMRLTVLLALLANFLPWPJAGAAPTADLVLTGVVAARAAVAILAVLL
ATFEVFLAKRLFRVPELLAAGSFLALLAULTAANFETVGA"
967..1629
/gene="hycp"
/gene="hycp"
967..1629
/gene="hycp"
/notes="Rv0085, (MTCY251.03), len: 220. hycP, similar to
sp|P77524|HYFE_ECOLI HYDROGENASE-4 COMPONENT E (216
aa)|FASTA scores: opt: 204 z-score: 272.1 E():1.2e-07;
25.5% identity in 216 aa overlap"
/codon_start=1
/transl_table=11
/product="hycp"
/protein_id="CAA98921.1"
/db_xref="GI:1405755"
/db_xref="SWISS-PROT:O10882"
/translation="MSNANFSILVDFAGGLVLASVLIWVRDLRAIVRLAWOGAAL
AAPIILGRINDRALIAVGIALLALRALVLPMLLARAAGAAQAQREATPLVNTASS
LLITAGLTATFATPQVNVLEPGVTINAVPAFAVLLIALFVMTIRLUHAYSQAGFL
MLDNGTAAATAFLLTAGVPLIVELGASLDVLFVAVIVIGLTLGRLRIIFGDADDLKREL
RP"
1629..3095
/gene="hycQ"
1629..3095
/gene="hycQ"
/notes="Rv0086, (MTCY251.04), oxidoreductase, len: 488,
hycQ, most similar to sp|P77437|HYFE_ECOLI HYDROGENASE-4
COMPONENT F(EEC 1.- (526 aa), FASTA scores: opt: 948
z-score: 1117.4 E(): 0: 35.9% identity in 493 aa overlap.
Also similar to E. coli d9087711 & NUOL_ECOLI P33607 nadh
dehydrogenase1 chain 1 (613 aa) FASTA scores, opt: 360,
z-score: 354.9, E():3.2e-13, (27.9% identity in 488 aa
overlap), and to NUON_ECOLI P33608, nadh dehydrogenase 1
chain n (425 aa), FASTAScores, opt: 375, z-score: 371.4,
E(): 3.9e-14, (25.0% identity in 432 aa overlap)"
/codon_start=1
/transl_table=11
/product="hycQ"
/protein_id="CAA98922.1"
/db_xref="GI:1405756"
/db_xref="SPTREMBL:O10883"
/translation="MTGLLLAIALPLAASLITGWRRTATITATLSATTVLACAV
AMGFMGSGAQFGLGLRADALTVMVLVIGVGTGLATAASIGYIDTELAHGHINGR
SARLYGLTPAFICAMVLAVCANNGIWIWAIEATITVITAFIVGHKRTALEATWKY
VVICSVGIAPFELGTVLLYFAARSGAAAAGALNDILAEHAAGLDPGVARIAGGLLL
IGYKAGAGLPFPHTLADAHQAAPAPVSAIMSGVLLAVAFSVLIHLRPLIDAVSGPAY
LRNGLLVGLTLLVAVLMLVTGVDVKPMLAYSSMEHGLIATAAAGTTIAAALL
HVLHAGIKGTVLFLAGLOQAADHSTAIADITGVMRRLIGVSFAVGLIIVLLGPPF
MTAFSELATARSANERLANWLGAAALLIITAGTALARNRSMRLTGLTTPAAGPATVTP
ATAAALMWGVVSAALGITAGPLADLLGLAASNVGLP"
3092..4570
/gene="hycE"
3092..4570
/gene="hycE"
/notes="Rv0087, (MTCY251.05), oxidoreductase, len:
492,hycE, similar to HYCE_ECOLI P16431 formate
hydrogenlyase subunit 5 (569 aa) FASTA scores, opt: 680,
z-score: 808.1, E(): 1.8e-38, (31.2% identity in 449 aa
overlap) and to NUOD_ECOLI P33600 nadh dehydrogenase 1
chain d, (407 aa) FASTA scores,opt: 245, z-score: 293.1,
E(): 8.9e-10, (24.5% identity in 368aa overlap)"
/codon_start=1
/transl_table=11
/product="hycE"
/protein_id="CAA98923.1"
/db_xref="GI:1405757"
/db_xref="SPTREMBL:O10884"
/translation="MMSASWLHRVSRERGLIATAEOLWADSFRLAIVAAHDDGDSLVR
VYLFAGYPPRRVELEYVVPADNPERSLAYLSFPAGREFERNADLYIRPVGHDPKPR
```

```
RLVRHAHPDWHPMRTDAGPAPEFTTDTGAPPELAVEGPGVYEIIPGVVHAGLIEPGHF
RESVAGETIVRLKARLWFWHRGTEKLPCHGRPATAVDAVLAERISGDT:SAAHALSHSLAI
EDALGIELPELVEHRLPALIVELERLYNHAADLGALANDCYSLANAAQKIRENLRLR
NAAVTGHRLRGALIRAGVALRALPDDELAALAVATLTLANSVYDORFACTA
VLHPDASALGCGIYVARASGLRSDTDEHPTIVLPITTEIGAPIGIIVLAKYIVRRDEF
AASAAIAQHIHVESHTGP IEYAATLHPVGAPSSGIGIVEGMRGTIVHKVF:IDVDGRITR
AKVDPSPFNPNPALIPVAMADTIVPDFPLANKSFNQSYAGNDL"
4605..5279
/gene="Rv0088"
4605..5279
/gene="Rv0088"
/notes="Rv0088, (MTCY251.06), len: 224. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0088"
/protein_id="CAA98924.1"
/db_xref="GI:1405758"
/db_xref="SWISS-PROT:Q10885"
/translation="MSVYKHAPSRLVRQTRSTVVKGRSGSLSNRRVTRTGDGLAVWG
GREYRAVKPCTPGIQPKGDMVTIVVDAGPGRVSRKSEVAAPAAEI:FAIVADPRRHR
ELDGSQTVRGNIKVPAKLVVGSKFSTKMKLFGLPYRITSRTALKPNELVWESHPLGHI
RWKEFESLSPTLTRVTETEDYHAAGAIAKGLKAYEMTGFAKSNAAI:FEATLAKLSQ
YAKGRA"
5436..6029
/gene="Rv0089"
5436..6029
/gene="Rv0089"
/notes="Rv0089, (MTCY251.07), len: 197. unknown, some
similarity to sp|P12999|BIOC_ECOLI BIOTIN SYNTHESIS
PROTEIN BIOC. (251 aa). FASTA scores: opt: 202 z-score:
262.0 E(): 4.5e-07; 39.0% identity in 118 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0089"
/protein_id="CAA98925.1"
/db_xref="GI:1405759"
/db_xref="SWISS-PROT:Q10886"
/translation="MDQPNWNIHYDALLDAMVPLGTQCVLIDVCCGIGLLAARLARRI
PYTAVDIDAEVILKRAQTFREANAPIRMLHADIMTAELPNAGFDUAVV:SNAAUHHIEDTR
TALSRLGLVTPGGT LAVVTEVTPSLRNLGHLTWSWACGMANRVKKGWHSAPIKWP
PPQTLHELRSVIRALLPGACIRLLYGRVLVTRAPV"
6147..6152
/notes="aaqaaq, potential rbs upstream of Rv0090"
6158..6928
/gene="Rv0090"
6158..6928
/gene="Rv0090"
/notes="Rv0090, (MTCY251.08), len: 256. Unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0090"
Query Match 38.9%; Score 176; DB 1; Length 38380;
Best Local Similarity 100.0%; Pred. No. 2.2e-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 278 cgcgggccaaccttgaagcgcgtgctagcgaacactgcccgcggcgagcaaacccq 337
Db 12127 CGCGGGCCACTTTGAAGCGGTGCTAGCCAAACTGCCGCCGCCGCCGCCGCCG 12068
QY 338 acgaccacaccccggtcatcgacacaccccgatcgccgcgcctcgccgcgacccc 397
Db 12067 ACGACCACACCCCGGTGTCGACACACCCCGATCGCGCGGCATCGACCGGACACC 12008
QY 398 gcgcgcaagcccaacgcaacacgacgagggctgctgcccgggctgcgcgcgtgac 453
Db 12007 GCAGCCAAAGCCCAACGCAACACGACGGGCTGCTGCGCGGCTGCGCGGCTGATC 11952
RESULT 14
MTY13E12 43401 bp DNA BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
DEFINITION
```

Wed Apr 3 08:33:13 2002

ACCESSION 295390 AL123456
VERSION 295390.1 GI:3261766
KEYWORDS
SOURCE
ORGANISM Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 43401)
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellion, S.,
Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature, 393 (6685), 537-544 (1998)
MEDLINE 98295987
REFERENCE 2 (bases 1 to 43401)
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2104370.
COMMENTS
Notes:
Details of M.tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of 79 genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-11bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES
source Location/Qualifiers
1..43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
misc_feature
1..1481
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y77"
78..1481
/gene="Rv3448"
78..164
misc_feature
/gene="Rv3448"
/note="PS00402 Binding-protein-dependent transport
systems inner membrane comp signature"
78..1481
CDS
/gene="Rv3448"
/note="Rv3448", (MTCY77.20), len: 467 aa. Unknown membrane
protein, contains PS00402 Binding-protein-dependent
transport systems inner membrane comp signature. Some
similarity to AL021930|MTV035.18 Mycobacterium
tuberculosis (472 aa) fasta score, opt: 429 z-score:
488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/product="hypothetical protein Rv3448"
/protein_id="CAB08736.1"
/db_xref="GI:3261767"
/db_xref="SPTREMBL:O33354"
/translation="MPTSDPGLRRVTVTVHAGAAQAVDLTPAAVPVAVILPSIVDILGDR
CASPAFAARVQISALCAPALPNATLQCCIRDGAVILVLRKSSAQIPPTSCDDVAEAV
AAALDITTAACQOKTTTLSCALAAATCITAGGGLMLVRLNLTGNVTKYSUATAGVVA
GIAALJFAVITACRTYRDPIAGLTSLVIAITFCGAVGLAVPGVKGIVHSVIAVMAA
TSVLAIRITCGGKITITAVACCACAVVAATLVCATVAAVPALGSLATIASGLLEVS
AKMAVLLAGISPRIPALNPADADALTTRDRLTRANRADAMITSLIAAFAASATIGA
IGTAVATHGIIHRSMGGLAAAVTGALLLLARSADYTRSLVFAICITTTATATFVA
ADNALEHCPWIAAI:TAMLAAMAVFI:GFVAPALSISPVYRTTLELLELALITAMVPLTA
WICGAYSAVRHIDL:TW"
611..>43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y13E12"
1478..2845
/gene="Rv3449"
1478..2845
/gene="Rv3449"
/note="Rv3449", (MTCY13E12.02), len: 455. Function:
probable precursor of serine protease. Has putative
peptide N-terminus and hydrophobic stretch at
C-terminus. Contains three signatures typical of subtilase
family: aspartic acid active site (PS00136), histidine
active site (PS00137), serine active site (PS00138). FASTA
results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E():
8e-07; (38.0% identity in 387 aa overlap), similar to
MTCY15F10.29 (45.7% identity in 451 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3449"
/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/translation="MTTSKTLRLVVSALATLSGLGTVAHAVSPPIIDERMLESAL
PAPRPTVQREVCYETVTAESGRAGRSQAQLADQVWRLTRGQRVAVIDTGV
RIIRLPKVVAGDGYFTGTGADCAHGTLVAGLIAAPDASDNFSCVAPDVLISI
RUSSEKFAIPGDPSTGVDVDMAKAVRTADLGCASVINISLACVPAAPADRAL
GAAALAYIVKNAVIAAAGTGGAAQCPGAVTRDSVTVAVSIAMVDYVLTGCS
VNAQHPSAFTLAFHVIDVAATGCAVTSISPGDGTGNRLRQHRSLPISOTSAAPV
VAGLAALIRARTLTITAGVVMQRIESTAHHPAGWDPLVGNCTVIALAAVSSDSIPQA
GLNLSFALVAVIVIRKSLPGSPDKKALHTATAGAACTLLALPATLATASRRIRPGRN
LADLP"
1757..1789
/gene="Rv3449"
/note="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862..1894
/gene="Rv3449"
/note="PS00137 Serine proteases, subtilase family,
histidine active site"
2456..2488
/gene="Rv3449"
/note="PS00138 Serine proteases, subtilase family, serine
active site"
complement(2810..4222)
/gene="Rv3450c"
complement(2810..4222)
/gene="Rv3450c"
/note="Rv3450c", (MTCY13E12.03c), len: 470. Function:
unknown but similar to Rv3895c MTCY15F10.17 (FASTA score:
1.6e-27; 36.4% identity in 475 aa overlap); possible
membrane spanning region near N-terminus, and to
Y14967|MLCB28.16 Mycobacterium leprae cosmid H628; (481
aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 32.9%
identity in 480 aa overlap, also similar to Rv3869
(MTV027.04)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"

```
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/translation="MPSPTATMLHVSQYRFLRLRIECALLFCDCVCAATGALRATTSL
ALGCVLAIVAMGCAFEVALLRPQALGAPIVMGRESGALYVRVDDVHVPVNLASAR
LIAATNANPQVSESEIGHTRKGLLPGAPQLLDQPLAGAESAWAICDSNNGSST
VYVGPDESSQAIVTEQMILVATESGPTLYLLGGRAVVDLADPAVYMWALRQGRV
PHVQAQSLNAPVAPRITAPRIIRGGGRASVGLPGFLVGGVVRITRASGDYVYVLED
QVIRIGOVAADLLRFQDSQSVNPTVAPDIRVAPIVNTLPVSAFDPDPTPVVGDSP
GRVTTLCVTTWTPAQGAARAFIAGSGPPVPLGSGVPTLAQADGRGPALDAVYLPQG
RSAYVAARSLSGGTGTRYLVITDTGVRFRIHDDVADHDLGLPTAAIPAPWPVLAFLPS
GPELSRANASVARDTVAPGP"
misc_feature
complement(4049..4081)
/gene="Rv3450c"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(4228..4231)
/note="possible RBS, GGAG, for Rv3450c"
4388..5131
/gene="Rv3451"
4388..5131
/gene="Rv3451"
/note="Rv3451"
/note="Rv3451, (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MYCTU_Q50664
probable cutinase cy339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/note="SWISS-PROT:O06318"
/transl_table=11
/translation="MGAGALITAVVLIALGAVWTPVAFADGCPDAEVTFAAGTGEPP
GIGRVGQAFVLSLRQQTGMEIGVYVNVYAAASRLQHLGGDGAIDAI SHIKSMASCPNT
KLVLGGYSOGATVIDI VAGVPLGSGISFGSPLPAAVADNVAAVFVGNPNRAGGSLSS
LSPLFGSKAIDLCPNTPDICHVGGNEFSGHIDYIPTYTTQAAAFVQRLRAGSVPH
LPGSVPLQSGVLQMPGTAAPAPESLHGR"
misc_feature
4706..4723
/gene="Rv3451"
/note="PS00155 Cutinase, serine active site, GGYSQ"
5165..5168
/note="possible RBS, GAGG, for Rv3452"
5178..5858
/gene="Rv3452"
5178..5858
/note="Rv3452"
/note="Rv3452, (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,
Query Match 38.9%; Score 176; DB 1; Length 43401;
Best Local Similarity 100.0%; Pred. No. 2.le-78;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 ccgcggccaccttgaagccgtgctagcacaactgcgcgcccccgcgcgaccccccgcg 337
Db 15904 CGCGGGCCACCTTTGAGCCGTGCTAGCCAAACTGGCCGCCCGCCGCGGACACCCCG 15963

QY 338 acgaccacccccggtcatcgacacaccccccgatgcgcgcgcccatcgaccgcgacacc 397
Db 15964 ACgACCACACCCCGTCATCGACACACCCCGGATCGCGGCCCGCCATCGACCGCACACC 16023

QY 398 cgagccaagcccaacgcaacacgagcgcgtgctgctgcgcggcgtgcgcgctgctgc 453
Db 16024 GCAGCCAGCCCAACGCAACGACGCGGCTGCTGCGCGGGCGCTGCGCGGCTGATC 16079

RESULT 15
AE007160
LOCUS AE007160 16384 bp DNA BCT 27-APR-2001
```

```
DEFINITION Mycobacterium tuberculosis CDC1551, section 246 of 280 of the
complete genome.
ACCESSION AE007160 AE000516
VERSION AE007160.1 GI:13883389
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
1 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..16384
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
96..839
/gene="MT3557"
96..839
/gene="MT3557"
/note="similar to SP:000298; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47897.1"
/db_xref="GI:13883390"
/translation="MGAGALITAVVLIALGAVWTLVAFADGCPDAEVTFAAGTGEPP
KLVLGGYSOGATVIDI VAGVPLGSGISFGSPLPAAVADNVAAVFVGNPNRAGGSLSS
LSPLFGSKAIDLCPNTPDICHVGGNEFSGHIDYIPTYTTQAAAFVQRLRAGSVPH
LPGSVPLQSGVLQMPGTAAPAPESLHGR"
871..1566
/gene="MT3559"
871..1566
/gene="MT3559"
/note="similar to SP:000298; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47898.1"
/db_xref="GI:13883391"
/translation="MRFIGVIRPQPHSGRWRAARLTLVAAFAAATLLTLLPAL
APPASGCPDAEVVFAAGTGEPPGLRGVQAFVSLRQTNKSICTYGVNYPANGDFL
AANDANDSHIQOMASACRATRLVLGGYSQGAVIDVTAAPLPGFPTPLPPAA
DDHTAATALFCNPSGRAGLMSALTPOFGSKTINLCNNGDPCISDCNHWRAHLGVYVG
MTNQAARFVASRI"
1624..1863
/gene="MT3560"
1624..1863
/gene="MT3560"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
```

gene	/product="hypothetical protein"	
	/protein_id="AAK47899.1"	
CDS	/db_xref="GI:13883392"	
	/translation="MTAADWHYAYHGLMLADPARQARRPTQPNNAHLSAECAAMNRN HTGREVGCGHANRTLGLTSTAIALRVADAGRHNQ"	
gene	1835..3523	
	/gene="MT3561"	
CDS	1835..3523	
	/gene="MT3561"	
gene	/note="identified by Glimmer2; putative"	
	/codon_start=1	
gene	/transl_table=11	
	/product="hypothetical protein"	
CDS	/protein_id="AAK47900.1"	
	/db_xref="GI:13883393"	
gene	/translation="MMPGVITNSSEPTAADHDIRITATRETLEDYTLRIAPRSYRWWP AVVGSALGIAYLADFAIGANVSETWGTANALCGIAIFALVVFYTGPLPAYAARYN IDLDTTRSGGYGVSVTVNIFATFTFFFALEGSIIMAOGLKLGHLPLWAGYACS TLIFPLVYGVKVLSQLQWTLPLWILMAAPFGVLYVSHPDLSIQFSPYAGKQDGHG GLSFGVLLAAGVCLSLIAQIAEQIDYLFEMPPRTPENANRWMTLLAGPCWAFGA TKOILGLFAYILMANIPGSSTIANQPVHOFQIYRTFVPCWLAULTLAVILVLSQIK INTNAYSGLAWNTSFTRLTKHYPRGVVFLGNLAIALILMEANWFDLNLILQCYA NCGMAWVAVASDIGNFYLLGLSPKTPFEFRGMLYAINPVFGFSLLLAAGLSIVTFP GGGAALQPSPLVAIVTALVMPPTLAAATKGXYLRTHDGDIDLPMDHGNPSAAV LTCHVCHQDFERPDMACOTGHGARVCSLCSTDKOAEHVLPLGLARAHIPGDQVP"	
	complement(3485..4378)	
CDS	/gene="MT3562"	
	complement(3485..4378)	
gene	/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U00096 PID:1788657; identified by sequence similarity; putative"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="tRNA pseudouridine synthase A"	
gene	/protein_id="AAK47901.1"	
	/db_xref="GI:13883394"	
CDS	/translation="MSLTRPPKSPQRPRIISGVVRLRLDIAVDCITDFAGWAAQVGQ RTVAGLDAAITIFTPVRLRAAGRTDAGVAGQVAVHDVPADALPNAYPRAGHV DPEFLPLLRGLRFLPADVRIIDITRAPAGDFARFSALRRHYVYRLSTAPYGVPEQQA RYITAMPREDLDAMTAASRLDMLHDFAFCFHREGATTIDILQRLDMSRAGTLVTA HVTADAFCSMWPSLVLCALLVGEHRRRTWCRELLTATCRSSDFAVAPAHGLTLIQ DYPFDQQLASRLNLYTRDVRSG"	
	complement(4323..4865)	
CDS	/gene="MT3563"	
	complement(4323..4865)	
gene	/note="similar to GB:J01685 SP:P02416 GB:X00766 PID:I47716 PID:42800; identified by sequence similarity; putative"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="ribosomal protein L17"	
gene	/protein_id="AAK47902.1"	
	/db_xref="GI:13883395"	
CDS	/translation="MPKPTKPRLGSSSHOKAILANLATSLFEHGRITTTPEKARAL RPYAEKLIHAKKALHNREVLKLRDKDVHTLFAEIGPFADRDGGYTRIKIEA RKGNAPMAVIELVREKTYTSEANRRARRVAAQAKAKAAMPTSEAKPAEEGDVW GASEPDAKAPPEEPTAPEP"	
	complement(4897..5940)	
CDS	/gene="MT3564"	
	complement(4897..5940)	
gene	/note="similar to SP:P72404; identified by sequence similarity; putative"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="DNA-directed RNA polymerase, alpha subunit"	
CDS	/protein_id="AAK47903.1"	
	/db_xref="GI:13883396"	
CDS	/translation="MLISQRPTLSEVDLTDNRSQFVIEPLEPGFCGYTLGNSLRRTLLS SIPGAAYTSIRIDGVLHEFTTVPGVKEDVTEIILNLKSLVYSEEDFPTMYLRKQGP GEVTAGDIVPAGVTHNPGMHITLNDKGKLEVLVVERGRGVYPAVQNRASGAIEG RIPVDSITYSVLKYVYKVDATRVQEDTFDKLLILDVETKNSISPRDALASAGKTLVEL FGLARELNVEAEGIEIGSPAEADHIAFALPIDDDLTLTVRSNCLAKREGVHTVELY	

us-09-785-904-2.olig.rge

Wed Apr 3 08:33:13 2002

Job time: 11239 sec

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 01:08:11 ; Search time 172.6 Seconds
(without alignments)
2250.106 Million cell updates/sec

Title: us-09-785-904-2

Perfect score: 453

Sequence: 1 gatcgcgagcgacatca.....ccgggctgcgcgcctaatc 453

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				N_Geneseq_1101.*	
1:	/SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*	2:	/SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*	3:	/SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*	5:	/SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*	6:	/SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*	8:	/SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*	9:	/SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*	11:	/SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*	12:	/SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*	14:	/SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*	15:	/SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*	17:	/SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*	18:	/SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*	20:	/SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*	21:	/SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*				

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB	ID	Description
1	176	38.9	650	17	AAT33658
2	176	38.9	650	19	AAV18649
3	176	38.9	1535	16	AAQ89200
4	176	38.9	1535	19	AAV33656
5	176	38.9	1535	19	AAV18647
6	176	38.9	12412	17	AAT33537
7	18	4.0	1413	22	AAH65816
8	18	4.0	1413	22	AAV1958
9	18	4.0	1536	21	AAH64886
10	18	4.0	1853	6	AAN50261
11	18	4.0	1853	10	AAN90631

12	18	4.0	8051	18	AAT72685	Sugar biosynthesis
13	18	4.0	8160	20	AAV5774	S.erythraea erythr
14	18	4.0	35026	21	AAA64890	Bordetella pertuss
15	18	4.0	349980	22	AAH68526	C glutamicum codin
16	17	3.8	281	21	AAH98508	Human colon cancer
17	17	3.8	300	20	AAV98072	Nucleotide sequenc
18	17	3.8	306	20	AAV98073	Nucleotide sequenc
19	17	3.8	441	22	AAI10473	Probe #406 for gen
20	17	3.8	441	22	AAI31455	Probe #413 used to
21	17	3.8	441	22	AAI31727	Probe #4609 used
22	17	3.8	555	22	AAI14923	Probe #13609 used
23	17	3.8	588	20	AAV98044	Nucleotide sequenc
24	17	3.8	742	13	AAQ20926	C10-E15 DNA fragme
25	17	3.8	932	13	AAQ20923	C10-E12 DNA fragme
26	17	3.8	957	20	AAV98046	Nucleotide sequenc
27	17	3.8	1077	22	AAH60764	Pseudomonas sp ABC
28	17	3.8	1188	22	AAH67681	C glutamicum codin
29	17	3.8	1410	17	AAV09251	SKK phosphotyrosin
30	17	3.8	1605	20	AAV98068	Nucleotide sequenc
31	17	3.8	2172	15	AAQ56930	Bacillus thuringie
32	17	3.8	2372	22	AAI59617	Human polynucleoti
33	17	3.8	2444	22	AAH60835	Pseudomonas sp exp
34	17	3.8	2592	22	AAH26409	Microbisporea thierm
35	17	3.8	2634	19	AAV22334	C glutamicum codin
36	17	3.8	3048	22	AAH65828	Nucleotide sequenc
37	17	3.8	3141	20	AAV98041	Human polynucleoti
38	17	3.8	4176	22	AAI57831	Human polynucleoti
39	17	3.8	4176	22	AAI57831	Human polynucleoti
40	17	3.8	5054	22	AAH07024	Pseudomonas putida
41	17	3.8	5544	22	AAH24891	Pimaricin biosynth
42	17	3.8	16382	16	AAH63074	IgG-FC binding pro
43	17	3.8	42235	20	AAV98035	Nucleotide sequenc
44	17	3.8	42235	20	AAV98035	Nucleotide sequenc
45	17	3.8	42235	20	AAV98035	Nucleotide sequenc

ALIGNMENTS

RESULT 1	
AAV33658	
ID	AAV33658 standard; DNA: 650 bp.
AC	AAV33658;
DT	22 Nov-1996 (first entry)
DE	M. tuberculosis macrophage survival gene.
KW	cellular uptake; cell entry; macrophage survival; vaccine;
KW	passive immunisation; gene therapy; ds.
OS	Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
XX	WO9626275-A1.
XX	29-AUG-1996.
XX	20-FEB-1996; 96wo-US02155.
XX	22-FEB-1995; 95US-0392210.
XX	(CORR) CORNELL RES FOUND INC.
XX	Riley LW;
XX	WPI; 1996-425086/42.
XX	P-PDB; AAW02303.
XX	DNA giving M. tuberculosis ability to enter mammalian cells
XX	and/or survive within macrophage(s), useful in vaccines to protect
XX	mammals against Mycobacterium tuberculosis infection

PS Claim 7; Page 45; 67pp; English.

XX A DNA molecule (AAT33658) codes for a protein (AAW02303) that confers

CC on Mycobacterium tuberculosis an ability to survive within

CC macrophages. It is a fragment of a larger molecule (AAT33656) obtd.

CC by ligating M. tuberculosis genomic DNA fragments into pBluescript

CC II vector and screening recombinant E. coli strains for HeLa cell-

CC invasive clones. It can be incorporated into a vector and used

CC for prodn. of recombinant macrophage survival protein, which is

CC useful in vaccines or for facilitating uptake of other materials,

CC e.g. therapeutic genes, into mammalian cells. A second DNA molecule

CC (AAT33657) codes for a protein (AAW02302) that confers on M.

CC tuberculosis the ability to enter mammalian cells.

XX

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 38.9%; Score 176; DB 17; Length 650;

Best Local Similarity 100.0%; Pred. No. 4.6e-77;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggcacatttgaagccgtgtagccaaactggccgccccggcgcgaccaaaccgcg 337

|||||

Db 221 cgcgggcacatttgaagccgtgtagccaaactggccgccccggcgcgaccaaaccgcg 280

Qy 338 acgacacacccccggtcatcgacacaccccccgatgcggccgcatcgacgcgacaccc 397

|||||

Db 281 acgacacacccccggtcatcgacacaccccccgatgcggccgcatcgacgcgacaccc 340

Qy 398 cgagccaaagcccaacgac 453

|||||

Db 341 cgagccaaagcccaacgac 396

RESULT 2

AAV18649

ID AAV18649 standard; DNA; 650 BP.

XX

AC AAV18649;

XX

DT 03-JUL-1998 (first entry)

XX

DE DNA for M. tuberculosis cellular uptake protein fragment.

XX

KW Cellular uptake protein; vaccine; infection; ds.

XX

OS Mycobacterium tuberculosis.

XX

FH Key Location/Qualifiers

FT CDS 1..649

FT /*tag= a

FT /note= "stop codon not given"

XX

PN W09805784-A1.

XX

PD 12-FEB-1998.

XX

PF 06-AUG-1997; 97WO-US13056.

XX

PR 07-AUG-1996; 96US-0689411.

XX

PA (CONN-) CONNAUGHT LAB LTD.

PA (CORR) CORNELL RES FOUND INC.

XX

PI Chong P, Riley LW;

XX

DR WPI; 1998-145620/13.

DR P-PSDB; AAW47543.

XX

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of

PT therapeutic agents e.g. antibiotics, also useful in vaccines

XX

PS Disclosure; Page 14; 82pp; English.

XX The present sequence encodes a Mycobacterium tuberculosis cellular

CC uptake protein fragment, which confers on M. tuberculosis an

CC ability to survive within macrophages.

CC The protein can be used in a vaccine to prevent M. tuberculosis

CC infection, and provide for the uptake in cells of, e.g.

CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies

CC raised against it can be used to treat mammals already exposed to

CC M. tuberculosis, to induce a passive immunity and prevent disease

CC occurrence.

XX

SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match 38.9%; Score 176; DB 19; Length 650;

Best Local Similarity 100.0%; Pred. No. 4.6e-77;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggcacatttgaagccgtgtagccaaactggccgccccggcgcgaccaaaccgcg 337

|||||

Db 221 cgcgggcacatttgaagccgtgtagccaaactggccgccccggcgcgaccaaaccgcg 280

Qy 338 acgacacacccccggtcatcgacacaccccccgatgcggccgcatcgacgcgacaccc 397

|||||

Db 281 acgacacacccccggtcatcgacacaccccccgatgcggccgcatcgacgcgacaccc 340

Qy 398 cgagccaaagcccaacgac 453

|||||

Db 341 cgagccaaagcccaacgac 396

RESULT 3

AAQ89200

ID AAQ89200 standard; DNA; 1535 BP.

XX

AC AAQ89200;

XX

DT 09-OCT-1995 (first entry)

XX

DE Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry

DE protein.

XX

KW Vaccine; tuberculosis; ss.

XX

OS Mycobacterium tuberculosis.

XX

FH Key Location/Qualifiers

FT CDS 1..1535

FT /*tag= a

XX

PN W09506726-A.

XX

PD 09-MAR-1995.

XX

PF 01-SEP-1994; 94WO-US09863.

XX

PR 02-SEP-1993; 93US-0118442.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Riley LW;

XX

DR WPI; 1995-115442/15.

DR P-PSDB; AAR71931.

XX

XX DNA encoding for cellular uptake of Mycobacterium tuberculosis

PT used to develop prods for vaccines, passive immunisation and

PT diagnosis and cellular uptake of other materials

XX

PS Claim 2; Page 9-11; 46pp; English.

XX

CC The isolated DNA molecule of the invention confers on M.

CC tuberculosis an ability to enter cells and to survive within

CC macrophages. It encodes a polypeptide having a mol. wt. of about 50-
CC 55 kDa, pI 5.2. The AA sequence represents a highly
CC hydrophilic protein with a hydrophobic region at its carboxy
CC terminus. It could be a secreted protein, a cytoplasmic protein, or
CC a surface protein with its carboxy terminus attached to the outer
CC membrane of the organism. The deduced AA sequence is in AAR71931.
XX
SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 38.9%; Score 176; DB 16; Length 1535;
Best Local Similarity 100.0%; Pred. No. 4.3e-77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcggccaccccttgaagcgtgtagcaaaactggccgcccccgcgagcaaaccccg 337
DB 1106 cgcggccaccccttgaagcgtgtagcaaaactggccgcccccgcgagcaaaccccg 1165
QY 338 acgacacaccccggtcgcacacaccccgatgcgcccccgccatcgaccgcgacacc 397
DB 1166 acgacacaccccggtcgcacacaccccgatgcgcccccgccatcgaccgcgacacc 1225
QY 398 gcagccaagcccaacgacgacgagggctgctggccgggtggcgcgctgac 453
DB 1226 gcagccaagcccaacgacgacgagggctgctggccgggtggcgcgctgac 1281

RESULT 4
AAT33656
ID AAT33656 standard; DNA; 1535 BP.

XX AAT33656;

XX 22-NOV-1996 (first entry)

XX M. tuberculosis cellular uptake gene region.

XX Cellular uptake; cell entry; macrophage; passive immunisation;
XX vaccine; gene therapy; ds.

XX Mycobacterium tuberculosis strain H37Ra (ATCC 25177).

XX Key Location/Qualifiers
XX 181..810
XX /tag= a
XX /label= ORF-1
XX /product= cell entry protein
XX 886..1535
XX /tag= b
XX /label= ORF-2
XX /product= macrophage survival protein

XX W09626275-A1.

XX 29-AUG-1996.

XX 20-FEB-1996; 96WO-US02155.

XX 22-FEB-1995; 95US-0392210.

XX (CORR) CORNELL RES FOUND INC.

XX Riley LW;

XX WPI; 1996-425086/42.

XX P-PSDB; AAW02301.

XX DNA giving M. tuberculosis ability to enter mammalian cells
XX and/or survive within macrophage(s), useful in vaccines to protect
XX mammals against Mycobacterium tuberculosis infection

XX Claim 2; Page 41; 67pp; English.

XX

CC A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an
CC ability to enter mammalian cells and to survive within macrophages.
CC The encoded protein sequence is given in AAW02301. The DNA was obtd.
CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II
CC vector and screening recombinant E. coli strains for Hela cell-
CC invasive clones. The DNA includes 2 separate coding regions (see
CC also AAT33657-58) coding for the cell entry (AAW02302) and macrophage
CC survival (AAW02303) proteins. It can be used to produce the cellular
CC uptake proteins used as vaccines or to facilitate uptake of other
CC materials, e.g. therapeutic genes.

XX
SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 38.9%; Score 176; DB 17; Length 1535;
Best Local Similarity 100.0%; Pred. No. 4.3e-77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcggccaccccttgaagcgtgtagcaaaactggccgcccccgcgagcaaaccccg 337
DB 1106 cgcggccaccccttgaagcgtgtagcaaaactggccgcccccgcgagcaaaccccg 1165
QY 338 acgacacaccccggtcgcacacaccccgatgcgcccccgccatcgaccgcgacacc 397
DB 1166 acgacacaccccggtcgcacacaccccgatgcgcccccgccatcgaccgcgacacc 1225
QY 398 gcagccaagcccaacgacgacgagggctgctggccgggtggcgcgctgac 453
DB 1226 gcagccaagcccaacgacgacgagggctgctggccgggtggcgcgctgac 1281

RESULT 5

AAV18647

ID AAV18647 standard; DNA; 1535 BP.

XX AAV18647;

XX 03-JUL-1998 (first entry)

XX DNA for M. tuberculosis cellular uptake protein fragment.

XX Cellular uptake protein; vaccine; infection; ds.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
XX 1..1534
XX /tag= a
XX /label= "Stop codon not given"

XX W09805784-A1.

XX 12-FEB-1998.

XX 06-AUG-1997; 97WO-US13056.

XX 07-AUG-1996; 96US-0689411.

XX (CONN-) CONNAUGHT LAB LTD.

XX (CORR) CORNELL RES FOUND INC.

XX Chong P, Riley LW;

XX WPI; 1998-145620/13.

XX P-PSDB; AAW47541.

XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
XX therapeutic agents e.g. antibiotics, also useful in vaccines

XX Disclosure; Pages 9-10; 82pp; English.

XX The present sequence encodes a Mycobacterium tuberculosis cellular
XX uptake protein fragment, which confers on M. tuberculosis an

CC ability to enter mammalian cells and to survive within macrophages.
 CC The protein can be used in a vaccine to prevent M. tuberculosis
 CC infection, and provide for the uptake in cells of, e.g.
 CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
 CC raised against it can be used to treat mammals already exposed to
 CC M. tuberculosis, to induce a passive immunity and prevent disease
 CC occurrence.
 CC
 XX Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;
 SQ

Query Match 38.9%; Score 176; DB 19; Length 1535;
 Best Local Similarity 100.0%; Pred. No. 4.3e-77;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggcacattgaagccgtctagcacaactggccgccccgcgcgacccaaccccg 337
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1106 cgcgggcacattgaagccgtctagcacaactggccgccccgcgcgacccaaccccg 1165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 338 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacccgacaccc 397
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1166 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacccgacaccc 1225
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 398 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacccgacaccc 453
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1226 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacccgacaccc 1281
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 6
 AAT33537
 ID AAT33537 standard; DNA; 12412 BP.
 XX
 AC AAT33537;
 DT 15-FEB-1998 (first entry)
 XX
 DE BCG deletion region 3 and flanking sequences.
 XX
 KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;
 KW mycobacteria; vaccine; infection; marker; ss.
 XX
 OS Mycobacterium bovis strain BCG.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1406..10673
 FT /tag= a
 FT /note= "BCG delta 1 deletion region"
 PN
 XX WO9625519-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 15-FEB-1996; 96WO-US01938.
 XX
 PR 17-FEB-1995; 95US-0390878.
 XX
 PA (PATH-) PATHOGENESIS CORP.
 XX
 PI Mahairas GG, Stover CK;
 XX
 DR WPI; 1996-393419/39.
 XX
 PT Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 PT bacterial infection
 XX
 PS Example 1; Fig 3; 66pp; English.
 CC
 CC This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta3. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT33535 and AAT33536) have also been detected.
 CC Identification involved screening a BCG cosmid library with a

CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.
 XX
 XX Sequence 12412 BP; 2146 A; 427 C; 3876 G; 1917 T; 200 other;
 SQ

Query Match 38.9%; Score 176; DB 17; Length 12412;
 Best Local Similarity 100.0%; Pred. No. 3.7e-77;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 278 cgcgggcacattgaagccgtctagcacaactggccgccccgcgcgacccaaccccg 337
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 737 cgcgggcacattgaagccgtctagcacaactggccgccccgcgcgacccaaccccg 796
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 338 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacccgacaccc 397
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 797 acgaccacaccccggtcatcgacacaccccgatgcccgcgcgcacccgacaccc 856
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 398 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacccgacaccc 454
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 857 gcagccaagcccaacgcaacacacaccccgatgcccgcgcgcacccgacaccc 912
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 7
 AAH65816/c
 ID AAH65816 standard; DNA; 1413 BP.
 XX
 AC AAH65816;
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 851.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790 A2
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Taleishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR P-PSDB; AAG90597.
 XX
 XX Novel polynucleotides derived from Corynebacterium, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 XX Claim 8; SEQ ID NO: 851; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium Corynebacterium glutamicum. These

Wed Apr 3 08:33:14 2002

are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 1413 BP; 259 A; 350 C; 395 G; 409 T; 0 other;

Query Match 4.0%; Score 18; DB 22; Length 1413;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 214 tcacctgagcaaccagc 231
|||||

Db 1007 TCACCTGAGCAACCAGC 990

RESULT 8
AAAF71958/c
ID AAF71958 standard; DNA; 1413 BP.

XX AAF71958;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:411.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

XX WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX (BADI) BASF AG.
XX Pompejus M, Kroeber B, Schroeder H, Zelder O, Haberhauer G;
PI WPI: 2001-137957/14.
DR P-PSDB; AAB79839.
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases.
XX Claim 3; Page: 759 /61; 1737pp; English.
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX Sequence 1413 BP; 264 A; 349 C; 394 G; 406 T; 0 other;
SQ

Query Match 4.0%; Score 18; DB 22; Length 1413;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 214 tcacctgagcaaccagc 231
|||||

Db 984 TCACCTGAGCAACCAGC 967

RESULT 9
AAA64886
ID AAA64886 standard; DNA; 1536 BP.
XX
AC AAA64886;

```

XX 02-FEB-2001 (first entry)
XX
XX Bordetella pertussis coding sequence # 2.
XX
XX Type III secretion system; virulence factor; pathogenicity island; ss.
XX
XX Bordetella pertussis.
XX
XX key Location/Qualifiers
XX CDS 1..1536
XX FT /*tag= a
XX ET /product= "Protein # 2"
XX
XX WO200037493-A2.
XX
XX 29-JUN-2000.
XX
XX 21-DEC-1999; 99WO-EP10297.
XX
XX 21-DEC-1998; 98GB-0028217.
XX
XX (ULBR ) UNIV LIBRE BRUXELLES.
XX
XX Bollen A, Fauconnier A, Godfroid E;
XX
XX WP.: 2000-452178/39.
XX
XX P-PSUB; AAB14148.
XX
XX Novel polypeptides derived from Bordetella pertussis, useful for
XX treating and diagnosing Bordetella infection -
XX
XX Example 2; Pages 154-156* 165pp; English.
XX
XX Bordetella pertussis possesses a type III secretion system. Type III
XX secretion systems allow bacteria to target virulence factors directly at
XX host cells. The present sequence is a coding sequence of B.
XX pertussis. No name has been specified for the present sequence. The
XX present sequence may be a housekeeping gene. A pathogenicity
XX island is a compact, distinct genetic unit carrying virulence genes. The
XX present sequence is located within a pathogenicity island (see AAA64890)
XX which also carries, a number of genes encoding proteins involved
XX in the type III secretion system of B. pertussis. See AAA64849-A64884
XX and AAB1411-B14146 for details of the coding sequences and proteins
XX identified in the pathogenicity island, of the present invention.
XX
XX Sequence 1536 BP; 248 A; 546 C; 560 G; 182 T; 0 other;
XX

```

Query Match 4.0%; Score 18; DB 21; Length 1536;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 ccgacggcgacctcaccg 187
|||
Db 593 ccgacggcgacctcaccg 610

RESULT 10

AA50261
ID AA50261 standard; DNA; 1853 BP.

XX
AC AAN50261:

14-JAN-1992 (first entry)

Plasmid pTrp1-35 DNA encoding the 2,5-diketogluconate-reductase gene.

KW 2,5-Diketogluconate-reductase; ascorbic acid; ss.

OS *Corynebacterium*.

XX	Key	Location/Qualifiers
FH		

FT	prcQuoter	229..234	
FT		/*tag= a	
FT	promoter	252..257	/label= E. coli trp promoter
FT		/*tag= b	
FT	RBS	/label= E. coli trp promoter	
FT		279..283	
FT	CDS	/*tag= c	
FT		296..1130	
FT		/*tag= d	
FT		/label= 2,5-DKG-reductase gene	
XX	EP132308-A.		
XX	30-JAN-1985.		
XX	25-JUN-1984;	84EP-0304277.	
XX	14-JUN-1984;	84US-0620585.	
XX	28-JUN-1983;	83US-0508409.	
XX	28-JUN-1983;	83US-0508410.	
XX	28-JUN-1983;	83US-0508628.	
XX	14-JUN-1984;	84US-0620651.	
XX	14-JUN-1984;	84US-0620652.	
XX	(GETH) GENENTECH INC.		
XX	Estell DA, Light DK, Rasteter WH, Lazarus RA, Miller JV;		
XX	WPT; 1985-026535/05.		
XX	P-PSDB; AAP50247.		
PT	New pure 2,5 di-keto-gluconic acid reductase - prepd. by		
PT	recombinant DNA methods for ascorbic acid prodn.		
XX	Disclosure; Fig 4; 44pp; English.		
XX	The 2,5-DKG-reductase gene is prepared by recombinant DNA tech		
CC	2,5-DKG-reductase is used to convert 2,5-DKG stereoselective		
CC	2-KLG, which is an intermediate in the production of ascorbi		
CC	(vitamin C).		
XX	Sequence 1853 BP; 327 A; 645 C; 587 G; 294 T; 0 other;		
SEQ			
Query Match	4.0%; Score 18; Len 6; Length 1853;		
Best Local Similarity	100.0%; Initd. No. 1;		
Matches 18; Conservative 0; Mismatches 0; Indels			
Qy	1/7 cyagctctccagcagccga 194		
Db	1039 cyagctctccagcagccga 1056		
RESULT 11			
AAN90631			
ID	AAN90631 standard; DNA; 1853 BP.		
XX			
AC	AAN90631;		
XX			
DT	22-JUN-1990 (first entry)		
XX			
DE	Sequence including 2,5-diketogluconic acid reductase gene an		
DE	control regions.		
XX			
KW	2,5-diketogluconic acid reductase; 2,5-DKG reductase;		
KW	2-keto-L-gluconic acid; 2-KLG; ascorbic acid; vitamin C; ss.		
XX			
OS	Corynebacterium sp. ATCC 31090.		
XX			
FH	Key	Location/Qualifiers	
FT	promoter	229..234	
FT		/*tag= a	

```

FT      promoter      /label=E. coli Trp promoter
FT      252..257      /*tag= b
FT      /label=E. coli Trp promoter.
FT      RBS           279..282
FT      /*tag= c
FT      mat_peptide   296..1129
FT      /*tag= d
FT      /product=2,5-DKG reductase
XX
XX      EP305608-A.
XX
XX      08-MAR-1989.
XX
XX      01-JAN-1987; 87EP-0202624.
XX
XX      28-JUN-1983; 83US-0508628.
XX      28-JUN-1983; 83US-0508409.
XX      28-JUN-1983; 83US-0508410.
XX      14-JUN-1984; 84US-0620651.
XX      14-JUN-1984; 84US-0620652.
XX      14-JUN-1984; 84US-0620585.
XX
XX      (GETH ) GENENTECH INC.
XX
XX      Light DR, Estell DA, Lazarus RA, Rastetter WH, Miller JV;
XX
XX      WPI; 1989-069875/10.
XX      P-PSDB; AAP94624.
XX
XX      Recombinant prodn. of 2,5-diketogluconic acid reductase - useful for
XX      converting 2,5-DKG stereoselectively into 2-ketogluconic acid, a
XX      precursor of vitamin C.
XX
XX      Disclosure; fig 4; 2lpp; English.
XX
XX      The enzyme encoded by the sequence, 2-KLG reductase is used to convert
XX      2,5-DKG into 2-KLG, a precursor of ascorbic acid.
XX      See also AAN94347 and AAN94348.
XX
XX      Sequence 1853 BP; 327 A; 644 C; 588 G; 294 T; 0 other;

Query Match      4.0%; Score 18; DB 10; Length 1853;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 cgacctcaccgcacccga 194
Db 1039 cgacctcaccgcacccga 1056

RESULT 12
AAT72685
ID AAT72685 standard; DNA; 8051 BP.
XX
XX      AAT72685;
XX
XX      18-SEP-1997 (first entry)
XX
XX      Sugar biosynthesis gene cluster.
XX
XX      Polyketide; glycosylation; eryBIV; eryBV; eryCVI; eryBVI; eryCIV;
XX      eryCV; eryBVII; L-mycarose; D-desosamine; antimicrobial; antibiotic;
XX      antifungal; fungicide; anticancer; cytostatic; anthelmintic; ss.
XX
XX      Saccharopolyspora erythraea.
XX
XX      Key      Location/Qualifiers
XX      CDS      80..1048
XX      /*tag= a
XX      /label= eryBIV
XX      1048..2295

```

```

FT      /*tag= b
FT      /label= eryBV
FT      2348..3061
FT      /*tag= c
FT      /label= eryCVI
FT      3214..4677
FT      /*tag= d
FT      /label= eryBVI
FT      4674..5879
FT      /*tag= e
FT      /label= eryCIV
FT      5917..7386
FT      /*tag= f
FT      /label= eryCV
FT      7415..7996
FT      /*tag= g
FT      /label= eryBVII
XX
XX      WO9723630-A2.
XX
XX      03-JUL-1997.
XX
XX      23-DEC-1996; 96WO-US20238.
XX
XX      21-DEC-1995; 95US-0576626.
XX
XX      (ABBO ) ABBOTT LAB.
XX
XX      Donadio S, Katz L, Staver MJ, Summers RG;
XX
XX      WPI; 1997-351066/32.
XX      P-PSDB; AAW19737-42.
XX
XX      New genes involved in sugar biosynthesis and attachment - used to
XX      generate polyketide antimicrobials etc. with altered pattern of
XX      glycosylation
XX
XX      Claim 1; Fig 4B; 85pp; English.
XX
XX      2. Polynucleotides (AAT72684 and AAT72685) represent 2 distinct clusters
XX      of genes encoding enzymes (AAW19734-36 and AAW19737-42) involved in
XX      sugar biosynthesis and attachment in Saccharopolyspora erythraea.
XX      The eryB genes are involved in the biosynthesis of L-mycarose and
XX      the eryC genes involved in the biosynthesis of D-desosamine.
XX      Novel glycosylation-modified polyketides are produced by
XX      selectively altering, inactivating or augmenting these eryB and/or
XX      eryC genes and introducing them into polyketide-producing
XX      microorganisms.
XX
XX      Sequence 8051 BP; 1225 A; 2898 C; 2765 G; 1163 T; 0 other;

Query Match      4.0%; Score 18; DB 18; Length 8051;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 419 acgacgggctgctgcccgc 436
Db 5311 acgacgggctgctgcccgc 5328

RESULT 13
AAAX25774
ID AAAX25774 standard; cDNA; 8160 BP.
XX
XX      AAAX25774;
XX
XX      08-JUN-1999 (first entry)
XX
XX      S.erythraea erythromycin-synthesis gene cluster eryAI-eryK.
XX      Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
XX      secondary metabolite; eryBII; eryCIII; eryCII; hybridisation; probe;

```

glycosylation; macrolactone; oleandomycin; ds.

Saccharopolyspora erythraea.

Key Location/Qualifiers

CDS 242..1210

FT /*tag= a

FT /label= ORF13

FT /gene= "eryBIV"

FT /product= "dTDP-keto-L-6-deoxyhexose-4-reductase"

FT 1210..2457

FT /*tag= b

FT /label= ORF14

FT /gene= "eryBV"

FT /product= "mycarosyltransferase"

FT 2510..3223

FT /*tag= c

FT /label= ORF15

FT /gene= "eryCVI"

FT /product= "dTDP-D-6-deoxyhexose-3-N-methyltransferase"

FT 3308..4840

FT /*tag= d

FT /label= ORF16

FT /gene= "eryBVI"

FT /product= "dTDP-4-keto-L-6-deoxyhexose-2,3-dehydratase"

FT 4837..6042

FT /*tag= e

FT /label= ORF17

FT /gene= "eryCIV"

FT /product= "dTDP-D-6-deoxyhexose-3,4-dehydratase"

FT 6080..7549

FT /*tag= f

FT /label= ORF18

FT /gene= "eryCV"

FT /product= "dTDP-D-4,6-dideoxyhexose-3,4-reductase"

FT 7578..8159

FT /*tag= g

FT /label= ORF19

FT /gene= "eryBVII"

FT /product= "dTDP-4-keto-D-6-deoxyhexose-3,5-epimerase"

XX WO9905283-A2.

PN 04-FEB-1999.

PD 98WO-FR01593.

XX 21-JUL-1998;

XX 12-JUN-1998; 98EP-0007411.

PR 25-JUL-1997; 97FR-0009458.

XX (HMRI) HOECHST MARION ROUSSEL.

XX Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC;

PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;

XX WPI: 1999-142938/12.

DR P-PSDB: AAW993389, AAW993390, AAW993391, AAW993392, AAW993393, AAW993394,

DR AAW993395.

XX New nucleic acid sequences encoding enzymes involved in macrolide

PT biosynthesis - useful for producing hybrid secondary metabolites,

PT particularly erythromycin analogues

XX Claim 8; Fig 3; 221pp; French.

XX This sequence represents the eryAI-eryK gene cluster from the

CC Gram-positive bacterium Saccharopolyspora erythraea which encodes

CC enzymes involved in the production of the macrolide antibiotic

CC erythromycin as a secondary metabolite. The erythromycin gene cluster

CC spans approximately 53 kb and contains at least 20 open reading frames

CC (ORF). This sequence contains the eryBIV, eryBV, eryCVI, eryBVI,

CC eryCIV, eryCV and eryBVI genes encoding ORFs 13, 14, 15, 16, 17, 18 and

CC 19 respectively. A DNA representing the eryG-eryAII1 region of the same

cluster is shown in AAX25772. The genes are used to produce hybrid

CC secondary metabolites in *S.erythraea*, i.e. erythromycin analogues which

CC may have improved properties or as hybridisation probes for isolating

CC homologous genes involved in glycosylation of macrolactones in

CC macrolide-producing strains (specifically oleandomycin-producing strains

CC of *Streptomyces antibioticus*).

XX Sequence 8160 BP; 1259 A; 2927 C; 2789 G; 1185 T; 0 other;

SQ

Query Match 4.0%; Score 18; DB 20; Length 8160;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 419 acgacgagctgcctgacgcg 436

|||||

Db 5474 acgacgagctgcctgacgcg 5491

RESULT 14

AAA64890/c

ID AAA64890 standard; DNA; 35026 BP.

XX AAA64890;

XX 02-FEB-2001 (first entry)

XX Bordetella pertussis pathogenicity island coding sequence.

XX

KW Bacterial infection; anti-bacterial; vaccine; whooping cough;

KW type III secretion system; virulence factor; pathogenicity island; ss.

XX Bordetella pertussis.

XX Key Location/Qualifiers

FT 711..2024

FT /*tag= a

FT /product= "Protein # 1 (AAB14147)"

FT complement (2055..3590)

FT /*tag= b

FT /product= "Protein # 2 (AAB14148)"

FT 4220..4696

FT /*tag= c

FT /product= "Protein # 3 (AAB14149)"

FT complement (4998..5948)

FT /*tag= d

FT /product= "Protein # 4 (AAB14150)"

FT 6160..6747

FT /*tag= e

FT /product= "Orf1 (AAB14132)"

FT complement (6783..7049)

FT /*tag= f

FT /product= "BscF (AAB14116)"

FT complement (7049..7338)

FT /*tag= g

FT /product= "BscH (AAB14115)"

FT complement (7379..8659)

FT /*tag= h

FT /product= "BscI (AAB14114)"

FT complement (8656..10755)

FT /*tag= i

FT /product= "BscD (AAB14111)"

FT complement (10752..11120)

FT /*tag= j

FT /product= "Orf2 (AAB14133)"

FT complement (11117..11527)

FT /*tag= k

FT /product= "Orf3 (AAB14134)"

FT complement (11532..11909)

FT /*tag= l

FT /product= "Orf4 (AAB14135)"

FT complement (11906..13003)

FT /*tag= m

```
FT /product= "BopN (AAB14131)"
FT 13002..13784
FT /*tag= n
FT /product= "Orf5 (AAB14136)"
FT 13806..14081
FT /*tag= o
FT /product= "Orf6 (AAB14137)"
FT 14097..14582
FT /*tag= p
FT /product= "BcrH (AAB14112)"
FT 14630..15571
FT /*tag= q
FT /product= "Orf7 (AAB14138)"
FT 15601..16803
FT /*tag= r
FT /product= "Orf8 (AAB14139)"
FT 16827..17288
FT /*tag= s
FT /product= "Orf9 (AAB14140)"
FT 17293..17814
FT /*tag= t
FT /product= "Orf10 (AAB14141)"
FT 17892..18218
FT /*tag= u
FT /product= "BscI (AAB14117)"
FT 18215..19039
FT /*tag= v
FT /product= "BscJ (AAB14118)"
FT 19032..19694
FT /*tag= w
FT /product= "BscK (AAB14119)"
FT 19664..20302
FT /*tag= x
FT /product= "BscL (AAB14120)"
FT 20307..21641
FT /*tag= y
FT /product= "BscN (AAB14121)"
FT 21641..22150
FT /*tag= z
FT /product= "BscO (AAB14122)"
FT 22147..22695
FT /*tag= aa
FT /product= "BscP (AAB14123)"
FT 22692..23771
FT /*tag= ab
FT /product= "BscQ (AAB14124)"
FT 23768..24439
FT /*tag= ac
FT /product= "BscR (AAB14125)"
FT 24445..24711
FT /*tag= ad
FT /product= "BscS (AAB14126)"
FT 24723..25523
FT /*tag= ae
FT /product= "BscT (AAB14127)"
FT 25520..26569
FT /*tag= af
FT /product= "BscU (AAB14128)"
FT 26566..26964
FT /*tag= ag
FT /product= "BscV (AAB14129)"
FT 26955..28757
FT /*tag= ah
FT /product= "BscC (AAB14113)"
FT complement (28778..29380)
FT /*tag= ai
FT /product= "BrpL (AAB14130)"
FT complement (29412..29591)
FT /*tag= aj
FT /product= "Orf11 (AAB14142)"
FT complement (29555..30529)
FT /*tag= ak
FT /product= "Orf12 (AAB14143)"
```

```
FT CDS 30631..31776
FT /*tag= al
FT /product= "Orf13 (AAB14144)"
FT complement (31773..33005)
FT /*tag= am
FT /product= "Orf14 (AAB14145)"
FT 32370..33014
FT /*tag= an
FT /product= "Orf15 (AAB14146)"
FT complement (33002..34852)
FT /*tag= ao
FT /product= "Protein # 5 (AAB14151)"
XX WO200037493-A2.
XX 29-JUN-2000.
XX 21-DEC-1999; 99WO-EP10297.
XX 21-DEC-1998; 98GB-0028217.
XX (ULBR ) UNIV LIBRE BRUXELLES.
XX Boilen A, Fauconnier A, Godfroid E;
XX WPI; 2000-452178/39.
XX P-PSDB; AAB14111, AAB14112, AAB14113, AAB14114, AAB14115, AAB14116,
XX AAB14117, AAB14118, AAB14119, AAB14120, AAB14121, AAB14122, AAB14123,
XX AAB14124, AAB14125, AAB14126, AAB14127, AAB14128, AAB14129, AAB14130,
XX AAB14131, AAB14132, AAB14133, AAB14134, AAB14135, AAB14136,
XX AAB14137, AAB14138, AAB14139, AAB14140, AAB14141, AAB14142, AAB14143,
XX AAB14144, AAB14145, AAB14146, AAB14147, AAB14148, AAB14149, AAB14150, AAB14151.
XX Novel polypeptides derived from Bordetella pertussis, useful for
XX treating and diagnosing Bordetella infection -
XX Example 2; Fig 5; 165pp; English.
XX Bordetella pertussis possesses a type III secretion system. Type III
XX secretion systems allow bacteria to target virulence factors directly at
XX host cells. A pathogenicity island is a compact, distinct genetic unit
XX carrying virulence genes. The present sequence encodes a pathogenicity island
XX from B. pertussis. The present sequence encodes a number of proteins
XX involved in the type III secretion system of B. pertussis i.e. Bordetella
XX pathogenicity protein. The proteins encoded by the present sequence may
XX be used to treat or diagnose B. pertussis infection, e.g. as a vaccine.
XX Whooping cough is a disease caused by infection by B. pertussis.
XX SQ Sequence 3502a BP; 5572 A; 11006 C; 12284 G; 6164 T; 0 other;
XX Query Match 4.0%; Score 18; DH 21; Length 35026;
XX Best Local Similarity 100.0%; Pred. No. 13;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 170 ccgacgcgcacrtccgcg 187
Dy 2998 CCGACGCGCACCCTCACC 2981
RESULT 15
AAH68526
ID AAH68526 standard; DNA; 349980 BP.
XX
XX AAH68526;
XX
XX 26-SEP-2001 (first entry)
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 7061.
XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
```

OS Corynebacterium glutamicum.
XX EPI108790-A2.
PN
XX
PD 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOWA) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI
XX WPI; 2001-376931/40.
DR
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Disclosure; SEQ ID NO: 7061; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP: 79274 A; 90638 C; 98727 G; 81341 T; 0 other;

Query Match 4.0%; Score 18; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 tcaccctgagcaaccagc 231
|
DB 201596 tcaccctgagcaaccagc 201613

Search completed: April 3, 2002, 02:03:22
Job time: 3311 sec


```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464.052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-464-052-1

Query Match          38.9%; Score 176; DB 3; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcgggcacacccgtgtagcgaacacaccccgatgcggcgccatcgacgcgacccc 337
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1106 CGCGGGCCACCTTTGAAGCGTGCTAGCAAACTGGCGCGCCCGCGGACCAACCCG 1165

QY 338 acgacacaccccggtcatcgacacacaccccgatgcggcgccatcgacgcgacccc 397
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1166 AGGACACACCCCGGTTCATCGACACACCCCGATGCGCGCGCCATCGACCGGACACC 1225

QY 398 gcagcagaccccaacgcacacgcgagcgtctgctgcccggcctgcgcgcgtgaltc 453
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1226 GCAGCCACGACCAACGACACACGAGCGGCTCTGGCGGGGCTGCGCGCGCTGATC 1281

RESULT 5
US-08-461-002-1
; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485H)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
```

```
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-461-002-1

Query Match          38.9%; Score 176; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 cgcgggcacacccgtgtagcgaacacaccccgatgcggcgccatcgacgcgacccc 337
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1106 CGCGGGCCACCTTTGAAGCGTGCTAGCAAACTGGCGCGCCCGCGGACCAACCCG 1165

QY 338 acgacacaccccggtcatcgacacacaccccgatgcggcgccatcgacgcgacccc 397
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1166 AGGACACACCCCGGTTCATCGACACACCCCGATGCGCGCGCCATCGACCGGACACC 1225

QY 398 gcagcagaccccaacgcacacgcgagcgtctgctgcccggcctgcgcgcgtgaltc 453
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1226 GCAGCCACGACCAACGACACACGAGCGGCTCTGGCGGGGCTGCGCGCGCTGATC 1281

RESULT 6
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.40
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 546
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-689-411-1

Query Match          38.9%; Score 176; DB 4; Length 1535;
```

```
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 1106 CGCGGGCCACCTTTGAAGCCGTGTAGCCAAACTGGCGGCCCGCGCGGCGGACCAACCCG 1165

QY 338 acgaccacaccccggtcatcgacaccccccgcgtagcgccgcccacgacccgacaccc 397
Db 1166 AGGACCACACCCCGTCATCGACACACCCCGGATGGCGCGCCCATCGACCGGACACCC 1225

QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgtgcccgggctgcgcgctgac 453
Db 1226 GCAGCCAAAGCCCAACGCAACACGACGCGGTGCTGGCGGGCTGCGCGCGCTGATC 1281

RESULT 7
PCT-US94-09863-1
; Sequence 1, Application PC/TUS9409863
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
; UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Michael L. Goldman
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy-disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: PCT/US94/09863
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman Mr., Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1000
; TELEFAX: (716)-263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1

Query Match 38.9%; Score 176; DB 5; Length 1535;
Best Local Similarity 100.0%; Pred. No. 1.1e-74;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 1106 CGCGGGCCACCTTTGAAGCCGTGTAGCCAAACTGGCGGCCCGCGCGGCGGACCAACCCG 1165

QY 338 acgaccacaccccggtcatcgacaccccccgcgtagcgccgcccacgacccgacaccc 397
Db 1166 AGGACCACACCCCGTCATCGACACACCCCGGATGGCGCGCCCATCGACCGGACACCC 1225

QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgtgcccgggctgcgcgctgac 453
Db 1226 GCAGCCAAAGCCCAACGCAACACGACGCGGTGCTGGCGGGCTGCGCGCGCTGATC 1281
```

```
RESULT 8
US-08-390-878-18
; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULANCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Hourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-18

Query Match 38.9%; Score 176; DB 1; Length 12412;
Best Local Similarity 100.0%; Pred. No. 9.5e-75;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 gcggggccacattgaagccgtgtgtagccaaactggccgccccggcgccgacaccccg 337
Db 747 CGCGGGCCACCTTTGAAGCCGTGTAGCCAAACTGGCGGCCCGCGGCGGACCAACCCG 796

QY 338 acgaccacaccccggtcatcgacaccccccgcgtagcgccgcccacgacccgacaccc 397
Db 797 AGGACCACACCCCGTCATCGACACACCCCGGATGGCGCGCCCATCGACCGGACACCC 856

QY 398 gcagcaagcccaacgcaacacacacgagcggtgtgtgcccgggctgcgcgctgac 453
Db 857 GCAGCCAAAGCCCAACGCAACACGACGCGGTGCTGGCGGGCTGCGCGCGCTGATC 912

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
```


COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/982,712
FILING DATE: 19921127
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-07-982-712-4

Query Match 3.8%; Score 17; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 gactggctacacccga-173
|||||

Db 7 GACTGGCTACACCCGA 23

RESULT 13
US-08-081-072-18
Sequence 18, Application US/08081072
Patent No. 5641654
GENERAL INFORMATION:
APPLICANT: No. 5641654oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS
TITLE OF INVENTION: DIAGNOSIS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dike, Bronstein, Roberts & Cushman
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-4280
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk, 3.50inch, 1.4Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh Classic
SOFTWARE: Microsoft Word Version 4.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,072
FILING DATE: June 22, 1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 742 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
US-08-081-072-18

Query Match 3.8%; Score 17; DB 1; Length 742;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 caatacgacggcatgctc 247
|||||

Db 358 CAATACGACGGCATGTC 374

RESULT 14
US-08-449-093A-18
Sequence 18, Application US/08449093A
Patent No. 5662906
GENERAL INFORMATION:
APPLICANT: No. 5662906oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS
TITLE OF INVENTION: DIAGNOSIS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-4280
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,093A
FILING DATE: May 24, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/081,072
FILING DATE: June 22, 1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/726,141
FILING DATE: July 8, 1991
CLASSIFICATION: 424
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 742 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
US-08-449-093A-18

Query Match 3.8%; Score 17; DB 1; Length 742;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 caatacgacggcatgctc 247
|||||

Db 358 CAATACGACGGCATGTC 374

RESULT 15
US-08-081-072-15
Sequence 15, Application US/08081072
Patent No. 5641654
GENERAL INFORMATION:
APPLICANT: No. 5641654oru MAKI, Kenjiro YAMAGUCHI, Ayumi
APPLICANT: TOYOSHIMA, and Michinori KOHARA
TITLE OF INVENTION: NON-A NON-B HEPATITIS-SPECIFIC
TITLE OF INVENTION: ANTIGEN AND ITS USE IN HEPATITIS

Wed Apr 3 08:33:14 2002

;; TITLE OF INVENTION: DIAGNOSIS
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dike, Bronstein, Roberts & Cushman
;; STREET: 130 Water Street
;; CITY: Boston
;; STATE: Massachusetts
;; COUNTRY: USA
;; ZIP: 02109-4280
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy Disk, 3.50inch, 1.4Mb storage
;; COMPUTER: Apple Macintosh
;; OPERATING SYSTEM: Macintosh Classic
;; SOFTWARE: Microsoft Word Version 4.0A
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/081,072
;; FILING DATE: June 22, 1993
;; CLASSIFICATION: 435
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 932 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to genomic RNA
;; US-08-081-072-15

Query Match 3.8%; Score 17; DB 1; Length 932;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 caatacagcggcatgtc-247
|||||
Db 424 CAATACGACGCATGTC 440

Search completed: April 3, 2002, 03:06:30
Job time: 8784 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:46:21 ; Search time 1557.61 Seconds
(without alignments)
3125.197 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacatca.....ccgggtgcgcgcgtgatc 453

Scoring table: OLIGO_NUC

Searched: Gapop 60.0 , Gapext 60.0
11351937 seqs, 5372889281 residues

Word size : 0
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlir:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	4.4	1100	11	BG285414
2	19	4.2	454	11	BF202807
3	19	4.2	458	11	BF924284
4	19	4.2	477	13	AO573457
5	19	4.2	494	10	AW287746
6	19	4.2	525	10	BE499331
7	19	4.2	563	10	BE490055
8	19	4.2	575	10	BE416426
9	19	4.2	581	10	AA949237
10	19	4.2	583	10	BF164631
11	19	4.2	598	11	BF164631
12	19	4.2	626	10	A1518445

13	19	4.2	650	11	BI168066
14	19	4.2	742	10	AA941556
15	19	4.2	761	13	AQ330146
16	19	4.2	802	10	AA403332
17	19	4.2	874	13	CNS01X7V
18	19	4.2	1163	13	BI10564
19	19	4.2	1273	13	AQ365308
20	18	4.0	256	11	BE850262
21	18	4.0	308	11	BF804832
22	18	4.0	355	11	BI135353
23	18	4.0	441	10	BE490643
24	18	4.0	446	10	AU091937
25	18	4.0	480	11	BF083249
26	18	4.0	483	11	BG560660
27	18	4.0	491	10	AW076894
28	18	4.0	519	11	BI142385
29	18	4.0	539	10	BE290933
30	18	4.0	554	13	AQ398116
31	18	4.0	577	10	AA654704
32	18	4.0	587	10	BE270509
33	18	4.0	638	10	AJ274323
34	18	4.0	670	11	BG810322
35	18	4.0	781	13	AQ848738
36	18	4.0	789	11	BG699432
37	18	4.0	796	11	BF268065
38	18	4.0	809	10	BE196478
39	18	4.0	856	13	CNS01XEP
40	18	4.0	857	11	BF026228
41	18	4.0	871	13	CNS04PLX
42	18	4.0	902	13	BH132519
43	18	4.0	911	11	BF784991
44	18	4.0	929	13	CNS01IWM
45	18	4.0	1009	10	AL534377

ALIGNMENTS

RESULT 1

BG285414 1100 bp mRNA EST 21-FEB-2001
LOCUS 602409782F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4539236 5',
DEFINITION mRNA sequence.

ACCESSION BG285414

VERSION BG285414.1 GI:13037347

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1. (bases 1 to 1100)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgepbbs-remail.nih.gov

Tissue Procurement: DCTD/DP

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10467 row: d column: 21

High quality sequence stop: 739.

Location/Qualifiers

1. .1100

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4539236"

/issue_type="NIH_MGC_91"

/lab_host="DH10B (phage-resistant)"

/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 255 a 330 c 280 g 232 t 3 others
ORIGIN

Query Match 4.4%; Score 20; DB 11; Length 1100;

Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 gacaccaccccgatgcgc 377
|||||
Db 967 GACACCACCCGATGCGC 986

RESULT 2

LOCUS BF202807 454 bp mRNA EST 06-NOV-2000
DEFINITION WHE1784_H02_004Zs wheat pre-anthesis spike cDNA library Triticum aestivum cDNA clone WHE1784_H02_004, mRNA sequence.
ACCESSION BF202807
VERSION BF202807.1 GI:111117549
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
1 (bases 1 to 454)
REFERENCE Anderson,O.D.,Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
CONTACT: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1..454
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1784_H02_004"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phuescript phagemids in the T7 Close lab (Choi, Close, Fenton) at the University of California, Riverside. plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

1..458
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone_lib "NIH255"
/dev_stage "Adult"
/note "Organ: not known tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini library was made by cloning products derived from OKSPES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 97 a 131 c 140 g 90 t
ORIGIN

BASE COUNT
ORIGIN

94 a 173 c 135 g 52 t

Query Match

4.2%; Score 19; DB 11; Length 454;

Best Local Similarity 100.0%; Pred. No. 69;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgccccagcggtc 153
|||||
Db 157 CCGCTACGCCACGCGGTC 175

RESULT 3

LOCUS BF924284 458 bp mRNA EST 19-JAN-2001
DEFINITION RC2-NT0225-271100-011-h09 NT0225 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF924284
VERSION BF924284.1 GI:12320172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 458)
REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&t2=RC2-NT0225-271100-011-h09&t3=2000-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 30
High quality sequence stop: 90.
Location/Qualifiers
1..458
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone_lib "NIH255"
/dev_stage "Adult"

TITLE
JOURNAL
MEDLINE
COMMENT

FEATURES
source

1..458
/organism "Homo sapiens"
/db_xref "taxon:9606"
/clone_lib "NIH255"
/dev_stage "Adult"
/note "Organ: not known tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini library was made by cloning products derived from OKSPES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 97 a 131 c 140 g 90 t
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 458;

Best Local Similarity 100.0%; Pred. No. 69;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 caaccacgacgagctgctg 432
|||||
Db 85 CAACCAGCGGCTGCTG 103

RESULT 4
AQ573457

LOCUS AQ573457 477 bp DNA GSS 02-JUN-1999
DEFINITION nbxb0081E24r CUG1 Rice BAC Library Oryza sativa genomic clone
ACCESSION AQ573457
VERSION AQ573457.1 GI:4973546
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 477)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GGAAACAGCTATGACCATC
 Class: BAC ends
 High quality sequence stop: 428.
FEATURES
 source
 1..477
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbxb0081E24r"
 /clone_lib="CUG1 Rice BAC Library"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBelogAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
BASE COUNT 84 a 137 c 118 g 138 t
ORIGIN
 Query Match 4.2%; Score 19; DB 13; Length 477;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 ttgcgcgccttttcgcc 45
 |||||
 Db 160 TTCGCGCCCTTTTCGCC 178
 |||||
RESULT 5
 AW287746/c
LOCUS LG1_271_H06_b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION AW287746

VERSION AW287746.2 GI:6859737
KEYWORDS EST.
SOURCE Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 494)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6677590.
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 401
 POLYA-No.
FEATURES
 Location/Qualifiers
 1..494
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Light Grown 1 (LG1)"
 /note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI ; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 82 a 202 c 148 g 62 t
ORIGIN
 Query Match 4.2%; Score 19; DB 10; Length 494;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 gccagggcgcgcgaagcga 94
 |||||
 Db 41 GCCAGGCCGCCGAGCCGA 23
 |||||
RESULT 6
 BE499341
LOCUS WHE0974_F08_L1525 Wheat pre-anthesis spike cDNA library Triticum
DEFINITION aestivum cDNA clone WHE0974_F08_L15, mRNA sequence.
ACCESSION BE499341
VERSION BE499341.1 GI:9697948
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 525)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.-C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - pre-anthesis spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773

Fax: 5105595818

Email: candersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

```

source
1. .525
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0973_F08_L15"
/tissue_type="Wheat pre-anthesis spike cDNA library"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site1: EcoRI; Site2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
```

BASE COUNT 111 a 200 c 153 g 61 t

ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 525;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||

Db 161 CCGCTAGCCCGCGGGTC 179
-.-

RESULT 7

BE490055 563 bp mRNA EST 31-JUL-2000

LOCUS WHE0364_A04_A08ZS Wheat cold-stressed seedling cDNA library

DEFINITION Triticum aestivum cDNA clone WHE0364_A04_A08, mRNA sequence.

ACCESSION BE490055

VERSION BE490055.1 GI:9609588

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 563)

AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
Seaton,C.L. and Tong,J.C.

TITLE The structure and function of the expressed portion of the wheat
genomes - Cold-stressed seedling cDNA library

JOURNAL Unpublished (2000)

COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: candersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

Location/Qualifiers
1. .563
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"

FEATURES

source

```

1. .575
/organism="Triticum aestivum"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone="MUG008_E09"
/tissue_type="young spikelets"
/dev_stage="Feekes' scale 6-7"
/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; M13 Reverse sequencing primer used. 1.2 kbp average
```

```

/clone="WHE0364_A04_A08"
/clone_lib="Wheat cold-stressed seedling cDNA library"
/tissue_type="Seedling"
/dev_stage="Five-day old seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site1: EcoRI; Site2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Five-day
old seedlings were transferred to 5 C cold room and kept
for 48 hr. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

BASE COUNT 110 a 206 c 163 g 84 t

ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 563;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgcccagcggtc 153
|||||

Db 154 CCGCTAGCCCGCGGGTC 172
|||||

RESULT 8

BE416425 575 bp mRNA EST 24-JUL-2000

LOCUS MUG008_E09R90628 ITEC MUG Wheat Spikelet Library Triticum aestivum

DEFINITION CDNA clone MUG008_E09, mRNA sequence.

ACCESSION BE416425

VERSION BE416425.1 GI:9414271

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 575)

AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Lantrigue,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pechioni,N., Qualseot,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sortells,M., Watkinson,M. and Wenzel,G.

TITLE International Triticale EST Cooperative (ITEC): Production of
Expressed Sequence Tags for species of the Triticeae

JOURNAL Unpublished (2000)

COMMENT Contact: Ogihara Y
Kihara Institute for Biological Research, Yokohama City University
Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

Location/Qualifiers
1. .575
/organism="Triticum aestivum"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone="MUG008_E09"
/tissue_type="young spikelets"
/dev_stage="Feekes' scale 6-7"
/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; M13 Reverse sequencing primer used. 1.2 kbp average

```

BASE COUNT      123 a   199 c   167 g   82 t   4 others
ORIGIN
    insert size."
Query Match      4.2%; Score 19; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgccagcggtc 153
|||||
Db 173 CCGTACGCCAGCGGGTC 191

RESULT 9
BE416426      581 bp      mRNA      24-JUL-2000
LOCUS      MUG008.E10R990628 ITEC MUG Wheat Spikelet Library Triticum aestivum
DEFINITION      cDNA clone MUG008.E10, mRNA sequence.
ACCESSION      BE416426
VERSION      BE416426.1 GI:9414272
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Triticum aestivum
REFERENCE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 581)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sortells, M., Warburton, M. and Wenzel, G.
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sortells, M., Warburton, M. and Wenzel, G.
TITLE      International Triticeae EST Cooperative (ITEC): Production of
JOURNAL      Expressed Sequence Tags for Species of the Triticeae
COMMENT      Unpublished (2000)
Contact: Ogihara Y
Kihara Institute for Biological Research, Yokohama City University
Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES
    source
    1..581
    /organism="Triticum aestivum"
    /cultivar="Norin 26"
    /db_xref="taxon:4565"
    /clone_lib="ITEC MUG Wheat Spikelet Library"
    /tissue_type="young spikelets"
    /dev_stage="Feekes' scale 6-7"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; M13 Reverse sequencing primer used. 1.2 kbp average
    insert size."
BASE COUNT      127 a   199 c   167 g   82 t
ORIGIN
    insert size."
Query Match      4.2%; Score 19; DB 10; Length 581;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 ccgctacgccagcggtc 153
|||||
Db 173 CCGTACGCCAGCGGGTC 191

RESULT 10
AA949237      583 bp      mRNA      EST
LOCUS

```

```

DEFINITION      LD28056.5prime LD Drosophila melanogaster embryo p0R2 Drosophila
melanogaster cDNA clone LD28056 5prime, mRNA sequence.
ACCESSION      AA949237
VERSION      AA949237.1 GI:3111146
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 583)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsunoda, G.,
Lewis, S. and Rubin, G.M.
BDGP/HIMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 280 row: E column: 8
High quality sequence stop: 400.
location/Qualifiers
    1..583
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="LD28056"
    /clone_lib="LD Drosophila melanogaster embryo p0R2"
    /sex="male and female"
    /dev_stage="0 to 24 hours mixed stage embryonic"
    /lab_host="XLI Blue"
    /note="Organ: embryo; Vector: p0R2; Site_1: EcoRI; Site_2:
    XhoI; Sized fractionated cDNAs were directly ligated into
    p0R2."
BASE COUNT      151 a   181 c   151 g   100 t
ORIGIN
    Query Match      4.2%; Score 19; DB 10; Length 583;
    Best Local Similarity 100.0%; Pred. No. 69;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccagcagcggtc 435
|||||
Db 543 CCACGACGGCTGTGGCC 561

REFSUIT 11
B1164631
LOCUS      B1164631
DEFINITION      B1164631 598 bp mRNA EST 09-JUL-2001
RE04343.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE04343 5 similar to CG4532;
Phan0004532 located on: X 6D1-6D2; 04/11/2001, mRNA sequence.
ACCESSION      B1164631
VERSION      B1164631.1 GI:14630437
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 598)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celnik, S. and Rubin
G.M.
BDGP/HIMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

```

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AF003438: arm: X 16347108,6646644]
estimated-cyto:6B1-6D3: 04/11/2001
Plate: RE.43 row: D column: 7
High quality sequence stop: 490.

FEATURES

source
Location/Qualifiers
1. .598
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE04343"
pFlc-1"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
BASE COUNT 156 a 185 c 157 g 99 t 1 others
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 598;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
Db 567 CCACGACGGGCTGCTGGCC \$85

RESULT 12

AI518445
LOCUS
DEFINITION
LD37958.5prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD37958 5prime, mRNA sequence.
ACCESSION AI518445
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; M.scomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE
JOURNAL
COMMENT
BDGP/RHMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 379 row: E column: 10
High quality sequence stop: 413.
Location/Qualifiers
1. .626
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD37958"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

FEATURES

source
Location/Qualifiers
1. .626
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD37958"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

BASE COUNT 165 a 192 c 162 g 107 t
ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 626;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
Db 563 CCACGACGGGCTGCTGGCC 581

RESULT 13

BI168066
LOCUS
DEFINITION
RE086607.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE086607 5 similar to CG4532: FBan0004532 located on: X 6D1-6D2; 04/11/2001, mRNA sequence.
ACCESSION BI168066
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
BDGP/RHMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003438: arm: X 16347108,6646644]
estimated-cyto:6B1-6D3: 04/11/2001
Plate: RE.86 row: A column: 7
High quality sequence stop: 515.
Location/Qualifiers
1. .650
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE086607"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
BASE COUNT 155 a 207 c 172 g 115 t 1 others
ORIGIN

FEATURES

source
Location/Qualifiers
1. .650
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE086607"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
BASE COUNT 155 a 207 c 172 g 115 t 1 others
ORIGIN

Query Match 4.2%; Score 19; DB 11; Length 650;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacgggtgctggcc 435
|||||
Db 589 CCACGACGGGCTGCTGGCC 607

RESULT 14
LOCUS AA941556 742 bp mRNA EST 23-APR-2001
DEFINITION LD25748.5prine LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD25748 5 similar to CG4532: FBan0004532 located on: X 6D1-6D2; 04/10/2001, mRNA sequence.

ACCESSION AA941556 GI:13767860
VERSION AA941556.2
KEYWORDS EST,
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 742)
AUTHORS Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G. M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT On May 1, 1998 this sequence version replaced gi:3101469.
Other ESTs: LD25748.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est6fruitfly.berkeley.edu
hit genomic AB003438; arm: X [6347108..6646644]
estimated-cyto: 6B1-6D3: 04/10/2001
plate: LD.257 row: D column: 12
High quality sequence stop: 605
POLYA=NO.

FEATURES
source
1..742
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="LD25748"
/clone="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

BASE COUNT 196 a 219 c 195 g 132 t
ORIGIN

Query Match 4.2%; Score 19; DB 10; Length 742;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 ccacgacggcgtactgccc 435
|||||
DB 571 CCACACGGGCTGCTGCC 589

RESULT 15
LOCUS AQ330146 761 bp DNA GSS 08-JAN-1999
DEFINITION nbxb0046h15f Rice BAC Library Oryza sativa genomic clone
AQ330146 nbxb0046h15f, DNA sequence.
ACCESSION AQ330146
VERSION AQ330146.1 GI:4121996
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 761)
Wing, R.A., and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genom

Unpublished (1998)
Contact: Wing RA
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence stop: 359.
Location/Qualifiers
1..761
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbxb0046h15f"
/clone_lib="CUGI Rice BAC library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBelouACL1; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 145 a 229 c 144 g 243 t
ORIGIN

Query Match 4.2%; Score 19; DB 13; Length 761;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 27 ttccgcgccttttttgcgc 45
|||||
DB 161 ttccgcgccttttttgcgc 174

Search completed: April 3, 2002, 01:30:31
Job time: 9850 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:44:13 ; Search time 2977.34 Seconds
(without alignments)
2510.033 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcgcgagcgacacatca.....ccggcgctgcgcgcgtgatc 453

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*, -
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_pl:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sv:*
- 13: gb_un:*
- 14: gb_vt:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_on:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_htgo_hum:*
- 31: em_htgo_inv:*
- 32: em_htgo_rod:*
- 33: em_htg_hum:*
- 34: em_htg_inv:*
- 35: em_htg_rod:*
- 36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	415.8	91.8	3453	1	MTU43540	U43540 Mycobacteri
2	410	90.5	1604	1	MBU35021	U35021 Mycobacteri
3	410	90.5	9281	1	MBDR351	U35017 Mycobacteri
C 4	410	90.5	9764	1	AE006921	AE006921 Mycobacte
C 5	410	90.5	10019	1	AF041819	AF041819 Mycobacte
C 6	410	90.5	17783	1	AE007028	AE007028 Mycobacte
C 7	408.4	90.2	1535	6	AR096713	AR096713 Sequence
C 8	408.4	90.2	1535	6	AR147694	AR147694 Sequence
C 9	408.4	90.2	32437	1	MTCY336	MTCY336
C 10	408.4	90.2	38380	1	MTCY251	MTCY251
C 11	408.4	90.2	43401	1	MTY14E12	MTY14E12
12	403.6	89.1	16384	1	AF007160	AF007160 Mycobacte
13	366.8	81.0	12412	6	186264	186264 Sequence 18
14	351	77.5	650	6	AR096715	AR096715 Sequence
15	351	77.5	650	6	AR147696	AR147696 Sequence
16	214.6	47.4	14179	1	AE007053	AE007053 Mycobacte
17	214.6	47.4	33100	1	MTCY949	MTCY949
C 18	205	45.3	34331	1	MTY165	MTY165
C 19	203.4	44.9	14953	1	AE006996	AE006996
C 20	189.2	41.8	40056	1	MSGY348	MSGY348
C 21	183.8	40.6	15019	1	AE007036	AE007036 Mycobacte
C 22	183.8	40.6	37432	1	MTY1125	MTY1125
C 23	137.4	30.3	15504	1	AE006995	AE006995 Mycobacte
C 24	135.8	30.0	22550	1	MTCY22G8	MTCY22G8
C 25	63.8	14.1	43254	1	MLCPRTN3	MLCPRTN3
C 26	63.8	14.1	312050	1	MLCPRTN3	MLCPRTN3
C 27	60.6	13.4	348450	1	MLEPRTN4	MLEPRTN4
C 28	59	13.0	819	11	CNS066WP	AL395663 T7 end of
29	57.2	12.6	64957	1	AB032367	AB032367 Streptomy
30	56	12.4	177883	2	AC046159	AC046159 Homo sapi
31	55.6	12.3	11096	1	AF275943	AF275943 Streptomy
C 32	55.6	12.3	12381	6	AX006889	AX006889 Sequence
C 33	55.6	12.3	64957	1	AB032367	AB032367 Streptomy
C 34	55.2	12.2	4231	8	SVHKGPG	X56010 Sorghum vul
C 35	54.2	12.0	197669	2	AC084064	AC084064 Homo sapi
C 36	54.2	12.0	265537	2	AC087228	AC087228 Mus muscu
C 37	54	11.9	56870	2	CEY39B6_3	Continuation (4 of
38	53.8	11.9	69173	2	AC025052	AC025052 Homo sapi
39	53.6	11.8	82024	2	AC023210	AC023210 Homo sapi
C 40	53.4	11.8	840	8	CNS0191G	AL111680 BoLyris
41	53.4	11.8	62649	2	AC022552	AC022552 Homo sapi
42	53.4	11.8	110737	2	AC011105	AC011105 Homo sapi
C 43	53.4	11.8	205914	2	AC027682	AC027682 Homo sapi
C 44	53.2	11.7	78220	2	AC023212	AC023212 Homo sapi
45	53.2	11.7	89994	2	AC021263	AC021263 Homo sapi

ALIGNMENTS

RESULT 1	MTU43540	MTU43540	3453 bp	DNA	BCT	14-AUG-1997
LOCUS	Mycobacterium tuberculosis rfbA, rhamnose biosynthesis protein (rfa), and rmlC genes, complete cds.					
DEFINITION	U43540					
ACCESSION	U43540.1	GI:2326948				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
	Mycobacterium tuberculosis.					
	Mycobacterium tuberculosis					
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;					
	Mycobacterium; Mycobacterium tuberculosis complex.					
REFERENCE	1 (sites)					
AUTHORS	Lee, J.					
TITLE	Rhamnose biosynthetic genes related to a novel repeated sequence of					
JOURNAL	Mycobacterium tuberculosis					
REFERENCE	2 (bases 1 to 3453)					
AUTHORS	Lee, T.					

causing a premature stop, and is not the result of a sequencing artifact; similar to SP:P16431 GB:X17506 PID:41684 PID:882614 GB:U00096; identified by sequence similarity; putative"

1449..2123
/gene="MT0096"
1449..2123
/gene="MT0096"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879143"
/translation="MSVYKKHPSRVRLRQTRSTVVKRSGSLSRVRVTGDLGLAVWG GREYRAVKPQTGIPQGDMMVTVVVDAGPGRVSRSEVAAAEALFAIADPRHR ELDSGTVRGNIKVPKLVGSKFTKMKILGFLPYRITSRVATKLNELVECSHPLGH RWRWFEESLPTLRVTETFDYHAAGAINKGLFEYEMTGFAKSNAAAGIATLAKLSDQ YACRA"

2280..2873
/gene="MT0098"
2280..2873
/gene="MT0098"
/note="similar to GP:145427; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAK44320.1"
/db_xref="GI:13879144"
/translation="MDQPMNATHYDALLDAMVLGTQCVDVCCGGLLAARLARRI PVTAVDIDAPVLRRAQTRFANAPIRWLHADIMTAELPNAGFDVAVSNAALHLEDR TALSLRGLGTPGGTLAVVTFVTPSLRNLGHLTSSWACGMANKVRKWEHSAPIKWP PPOTLHELHSHVRLALPGACIRRLLYGRVLVTRAPV"

3002..3772
/gene="MT0099"
3002..3772
/gene="MT0099"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44321.1"
/db_xref="GI:13879145"
/translation="MAKNQNRINRWELITCGLGHVYAPDAAALAAKRASTGLGE WKLRCGDFALGGPOGRGAPEDAPLIMRGKALROALIIIRALGVERLVALVLAAW AVWEFRGARGAQTALDRDLPVLRAGFKVDQMTVIHALEKALAKPSTLALITCMLA AYAVLOAVEGVGLWLLKRWGEYFAVVATSIPLFLEVHDIAGIITTRVVTTSINVAAV VYLLSKRLFGVRGRKAYDVERRGQLDLLEKAMLT"

3791..4162
/gene="MT0099.1"
3791..4162
/gene="MT0099.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44322.1"
/db_xref="GI:13879146"
/translation="MRGLRVYSKVTSAFFALTTPARRARKAARVNLDOVAOCRRTDEG PTLQCHQPGFSALPAAWSRQSRVPATHCRPPCAPCAASALTALCAEAWSWVE VRPAPRAATFSQSRAARESR"

4206..4973
/gene="MT0100"
4206..4973
/gene="MT0100"
/note="similar to GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase, putative"

/protein_id="AAK44323.1"
/db_xref="GI:13879147"
/translation="MAVTGVICAIPQELAYLGVLDAKROQVAQILFDSQDLAHR VLLAAGMGKVNLTGLTATLLADRGCTITFTSVAGGLDELICGDIVIADRVQIDF GLITDERLPKVPCHIPFIEPTERLGYPVDPVIDRVKRLDGLTAPLSTAAAGGR OPRIYGTILTCDOYLHCFTNRRLHIELGGMAVEGGAQAQICASFPIIPWILVIAL SLDLAGDSGVDFNRFVGEVAASARVILRLLPVLTAQ"

5105..7390
/gene="MT0101"
5105..7390
/gene="MT0101"
/note="similar to SP:p46840 PID:559912; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK44324.1"
/db_xref="GI:13879148"
/translation="MTAVTGEHSHASVQRIQLRISGMSCSACAHVESTLNKLPGVRA AVNGTRVATIDTSEAVIAAALCOAVRAGYQADLCYDDGSRASDPPADHARQLIRL AIAVLFPVADLIVMEGVIPATFTGNWQVLSALPLPVVTAAMPFHRVAMRNARHH VGRFEAKAKSOAGSALKALAAKAKAVALLPDGSEMWIPADFLKEQRFVVRPGQI AASMETLISVGITATATNSLYTVEGNISPIERSGIMQALLSDAIFYEVAAGVTFVL VAAIGLAVDGSAAVIAQRIADRISSVFPVAVIVIAALTAGMLIAGCPDRAVSAAL AGMKLVFQAQAKADIAQRIADRISSVFPVAVIVIAALTAGMLIAGCPDRAVSAAL AVIVIAACALCIIATITAMVASGQAQGLIFLKGYSLEATRAVDTTFDKTGTITTT CBLQVSATTAAGWEADQVLAIAATVEAASRHSVALIAAATTRDAVTFDRAIPGRG VSGTVSGRAVRVGPSTWISGSSCHPNMRAARHAEISLETAFVEVDGECGVTAAD AVKTSARIAVAALADKGLRTMLITGDNPESAATAVATRGIDEVIADILPEKQVDVIEQ LDRGHVAVMGVGNINIKIPALIRADLMAIGCTDVAIGAADIILVKHULDVVPIALD LARATMTVTKLNNWAFGYNIAIPVAAGLLNPLVAGAAAFSFFEVVSNILRKPF GRYPGCGTGVGGQMTAPSSA"

complement(7337..8185)
/gene="MT0102"
complement(7337..8185)
/gene="MT0102"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44325.1"
/db_xref="GI:13879149"
/translation="MIAOATTAGSFNIIHASTVLOGCRGVPAAWSEPACAIHHCATI DMCDEVARFALSARLIGKRAIVPSARVDEHLGECGSACRAMFTQVASQADLRLAES RPVVPVGRGLGIRAPKPKHSPMTRRRWALLCVGIAOTALGTQVGFGLDVLTHQHT GACTHLLINESTWSIALCVIHYGAALWPSAAGLAGVLTAFVAILTGYVIVDALSCAV STTRIILHLPVVICAVLAIMWRSASGRPRPDVANAEPDILVLPDNASKRRRKHLPW TDKSAA"

complement(8232..9093)
/gene="MT0103"
complement(8232..9093)
/gene="MT0103"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44326.1"
/db_xref="GI:13879150"
/translation="MDWLHPGDLTDTFERAKRGITLSNOQYDMSRLSGYLIPQARA TFEAVLAKIAPGATNPDDHTPVDTTPDAAADIDROTQSOQRNHDGILAGLALIAS KGLQHGCLPVSIVVTITLDTOTGACKGFTGGTLLPNADVIKMTSHAHYSPASGR YPQAIIPDHGPTLALYHTRKLASPAQRIHLPANDRGCTKCCDAPAYHSDAHVTAWTS TGRDITELTLCGPDNRLAEKGTTHKNTHTGHTEWLPPLPHLDHGQPRNTFIHUPERF LHNQDDDDKPD"

complement(8956..9623)
/gene="MT0104"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; identified by Glimmer2; putative; conserved hypothetical protein, authentic frameshift"


```
CDS      7468..9051
        /gene="MT1631"
        /note="similar to GP:608530; identified by sequence
        similarity: putative"
        /codon_start=1
        /transl_table=11
        /product="L-aspartate oxidase"
        /protein_id="AAK45899.1"
        /db_xref="GI:13881260"
        /translation="MAGPAWRDAADVIVIGTVAGVLAALAAADRGRSVVLSKAQT
        HVTATHYAQQGIAVILPDDSDVDAHVDTLAAGAGLCPDPAVYSIVADGYRAVTDLV
        GAGRLDSEVPGRWALTREGHSRRRIHVAGGDATGAEVORALQDAAGMLDIRTHVA
        LRVLHDGTAVTGLLVVRPDGCGIISAPSVTLTGLGLHLYSATNPAGSTGGLALGL
        WAGVAYDLDFIOFHPTMLPAGRGRRLPTAIRGKSAIIVDRGNCNITAGVHPMG
        DLAPDVAAIDARLAKATGDCPCYILDAKIEGFASRFPVTATSCRAAGIDVVRQPI
        VVPAHYSCGIVDVTGTELLGLYAAGEVARTLGHANRLASLLLEGLVVGGRAG
        KAAAHAAAGRSRATVSSATWPEPISYTDLDGLQRAMSRDASMYRAAAGLHRLCDS
        LSGAQRVDVACRRDFEDVALTLVAQSVTAALARTESRGCHRAEYPTVPTVPEARSIV
        VRGADDANAVCVQALVAVC"
        gene      9021..9908
CDS      /gene="MT1632"
        /gene="MT1632"
        /note="similar to GB:L20833 SP:F30011 GB:L28105 PID:310262
        PID:456041; identified by sequence similarity: putative"
        /codon_start=1
        /transl_table=11

Query Match      90.5%; Score 410; DB 1; Length 17783;
Best Local Similarity 98.0%; Pred. No. 1.7e-56;
Matches 447; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

QY 1 gatcgcgagcgcaoaetcaagtgattcgccctttttcgccacccgacccgagcgcg 60
DB 1125 GATCGCGAGCGCACATCAAGTATCGCGCCCTTTTTCGCCACCTGCCCGCGCG 1066

QY 61 tggatgtgtccaccgagcgccggaagcgagcgacgtcgcgcaaggc--ctcaatalc 118
DB 1065 TGGATGTGTCCACCAGCGCGCGGCGGAAAGCGACCTGGCGCGCAAGCGCTCAATATC 1006

QY 119 gtcccgagcggtgcccgcctacgcccagcggtgatgactggtctacaccccgagcgc 178
DB 1005 GTCCCGAGCAGGTGCGCCGCTACGCCACGCGGTATGACTGGCTACACCCCGCGCG 946

QY 179 acctaccgacacgacgcccgcgcaaacgc-gcataccctgagcaaccagcaatc 237
DB 945 ACCTACCGACACCGAACCGCGCCGCGCAACCGCGCATCACCTTGAGCAATATG 886

QY 238 acggcatgtcacggctaaagtgtgctacgtcgcgcggtgatgactggtctacaccccgagcgc 297
DB 885 ACGGCATGTACGGCTAAGTGCTACCTGACCCGCCAAG-CGCGGGCCACCTTTGAAGCC 827

QY 298 gtgctagcaaaactggcgccgcccgcgcaaacgc-gcataccctgagcaaccagcgtc 357
DB 826 GTGCTAGCAAAACTGGCGCCGCGCGCGACCAACCCCGACGACACCCCGGTCAATC 767

QY 358 gacaccaccccgatgcgcgcccatcgacgcgacaccccgacgcaagcccaagcaac 417
DB 766 GACACCACCCCGATGCGCGCGCATCGACCGGACACCCCGACGCAAGCCCAACGCAAC 707

QY 418 cagcagcggtgctgcccgggctgcccgcgtgctgac 453
DB 706 CACGACGGGCTGCTGGCGGGCTGCGCGCGCTGATC 671

RESULT 7
AR096713
LOCUS      AR096713      1535 bp      DNA      PAT      08-SEP-2000
DEFINITION      Sequence 1 from patent US 6008201.
ACCESSION      AR096713
VERSION      AR096713.1
KEYWORDS      AR096713.1 GI:10025749
SOURCE      Unknown.

QY 1 gatcgcgagcgcaoaetcaagtgattcgccctttttcgccacccgacccgagcgcg 60
DB 827 GATCGCGAGCGCACATCAAGTATCGCGCCCTTTTTCGCCACCTGCCCGCGCG 886

QY 61 tggatgtgtccaccgagcgccggaagcgagcgacgtcgcgcaaggc--ctcaatalc 118
DB 887 TGGATGTGTCCACCAGCGCGCGGCGGAAAGCGACCTGGCGCGCAAGCGCTCAATATC 946

QY 119 gtcccgagcggtgcccgcctacgcccagcggtgatgactggtctacaccccgagcgc 178
DB 947 GTCCCGAGCAGGTGCGCCGCTACGCCACGCGGTATGACTGGCTACACCCCGCGCG 1006

QY 179 acctaccgacacgacgcccgcgcaaacgc-gcataccctgagcaaccagcaatc 237
DB 1007 ACCTACCGACACCGAACCGCGCCGCGCAACCGCGCATCACCTTGAGCAATATC 1066

QY 238 acggcatgtcacggctaaagtgtgctacgtcgcgcggtgatgactggtctacaccccgagcgc 297
DB 1067 ACGGCATGTACGGCTAAGTGCTACCTGACCCGCCAAG-CGCGGGCCACCTTTGAAGCC 1125

QY 298 gtgctagcaaaactggcgccgcccgcgcaaacgc-gcataccctgagcaaccagcgtc 357
DB 1126 GTGCTAGCAAAACTGGCGCCGCGCGCGACCAACCCCGACGACACCCCGGTCAATC 1185

QY 358 gacaccaccccgatgcgcgcccatcgacgcgacaccccgacgcaagcccaagcaac 417
DB 1186 GACACCACCCCGATGCGCGCGCATCGACCGGACACCCCGACGCAAGCCCAACGCAAC 1245

QY 418 cagcagcggtgctgcccgggctgcccgcgtgctgac 453
DB 1246 CACGACGGGCTGCTGGCGGGCTGCGCGCGCTGATC 123;

ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 1535)
AUTHORS      Riley, L.W. and Chong, P.
TITLE      DNA molecule fragments encoding for cellular uptake of
JOURNAL      Mycobacterium tuberculosis and uses thereof
FEATURES      Patent: US 6224881-A 1 01-MAY-2001;
              Location/Qualifiers
              Source
              1..1535
              /organism="unknown"
              BASE COUNT      297 a      544 c      458 g      236 t
              ORIGIN
```

Query Match 90.2%; Score 408.4; DB 6; Length 1535; Best Local Similarity 97.8%; Pred. No. 8.3e-56; Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;																				
Qy	1	g	atcg	g	g	g	g	g	g											
Db	827	G	A	T	C	G	G	G	G											
Qy	61	t	g	a	t	g	t	t	c											
Db	887	T	G	A	T	T	C	C	C											
Qy	119	g	t	c	c	a	g	a	g											
Db	947	G	T	C	C	A	G	A	G											
Qy	179	a	c	c	t	a	c	c	a											
Db	1007	A	C	C	T	A	C	C	A											
Qy	238	a	c	g	g	c	a	t	g											
Db	1067	A	C	G	G	C	A	T	G											
Qy	298	g	t	g	t	a	g	t	a											
Db	1126	G	T	G	T	A	G	T	A											
Qy	358	g	a	c	a	c	c	c	a											
Db	1186	G	A	C	A	C	C	C	A											
Qy	418	c	a	c	a	c	c	c	a											
Db	1246	C	A	C	A	C	C	C	A											
RESULT	9																			
LOCUS	MTCY336	32437 bp	DNA	BCT	03-AUG-2001															
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.																			
ACCESSION	295586	AL123456																		
VERSION	295586.1	GI:3261785																		
KEYWORDS	Mycobacterium tuberculosis H37Rv.																			
SOURCE	Mycobacterium tuberculosis H37Rv																			
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.																			
REFERENCE	1 (bases 1 to 32437)																			
AUTHORS	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.																			
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence																			
JOURNAL	Nature. 393 (6685), 537-544 (1998)																			
MEDLINE	98295987																			
REFERENCE	2 (bases 1 to 32437)																			
AUTHORS	Parkhill, J.																			
TITLE	Direct Submission																			
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2117233.																			
COMMENT	Notes:																			

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of 78 genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers

1..32437
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..273
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y48"
33..32091
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y336"
complement(39..2228)
/gene="RV1565c"
complement(39..2228)
/gene="RV1565c"
/note="RV1565c" (MTCY336.38), len: 729. Unknown membrane protein, some similarity to O05402 HYPOTHETICAL 72.2 KD PROTEIN from B. subtilis (634 aa) opt: 384 E(): 4.8e-17; 29.1% identity in 378 aa overlap and to hypothetical protein in H. influenzae. N-terminal half hydrophobic.
FASTA results, Y992_HAEIN P43993 h10392 (245 aa) opt: 265; E(): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse score is 0.930"
/codon_start=1
/transl_table=1
/product="hypothetical protein RV1565c"
/protein_id="CAH09072.1"
/db_xref="GI:2117272"
/translation="M.TLSIPRPPALTPPALPPVTMGTRTTGPFYRHDLDLGRGVAIA
LVVLAALILITAIQIQTHWEAFANQSLASGLYUNWELASTVSNYLKAGAVSPLO
HIVMSVQGGYIAFILLVAGCAYLLRRFGKPRAPYLRMTFVLLSTLTASFIVAI
VAHHAYQATAYNTFAKAWELLAGALVGVVPIHVPWPMWLKTAATAALAAIISGAL
HVCVFPGIWIAPVGCATMILACANRQGHPTDRDLPINRLIATAPILVAGAMA
KSYLVHMPILLIPWISYTGHRHANFVGEAVLLVGLLAYLTRIVEDPLRYKAPAGV
KSPAAVPIIPWIRLIRPTITVLGVSVALGLVALTSFTWREHVIQVRAAGKELISLS
SRDPGARALIDHVRPKLHMRPTVLEVRDLPTSKDGCISDFVNPATINCTYGOVD
APHTIALAGSHAELHLLDILGRHHFKVYVYLMKGCPLSTEVEPLIMGNAPVPO
CHQWQAAMAKLVADHIDPDYVFTSTPWNTPKGVMPATYVGVQWTFADNNIPVLAMR
ITPWLKVDGQFFIPADCLAKGNGPQSGIARSKVLVDNRNPTLDFVARFPLKPLDMSD
AICRTDTFCRAVEGNVLVYRDSHHLLTPTYMTMTSELGRQIAANTDMW"
complement(2327..3019)
/gene="RV1566c"
complement(2327..3019)
/note="RV1566c" (MTCY336.37), len: 230. Function: unknown
probably exported has QQAPV repeats at C-terminus, similar to hypothetical M. leprae protein, to putative invasins 1,2(007390, 007391) from M. avium and slightly similar to C-terminus of Listeria invasion-associated protein p60 precursor P60_LisMO P21171. FASTA results, Q49634 COSMID B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa overlap. TBparse score is 0.890"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1566c"
/protein_id="CAB09071.1"
/db_xref="GI:2117271"
/db_xref="SPTREMBL:O06624"
/translacion="MKRSKSGSFAIGLAMLPVMAAPGLAAADPATRPVDYQOIID
VITARGLSORGVPSWAGGSGISGTRGTGCTGNTVGEDASGLIOYVAGAGLKLPRSS
GOMYKGVKVLPOQARKDLIFYGPECTQSVALYLGKQMLEVGVQVVOVSPVRTNGMT
PILRVLGTQPTPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAPVQAP
VQPPFGTARSR"
complement(3030..3036)
/notes="possible RBS, AAGGAGG, for Rv1566c"
complement(3258..3263)
/notes="ASNI site: ATTAAT; probably linking fragments
B3/G6"
complement(3259..3543)
/genes="Rv1567c"
complement(3259..3543)
/genes="Rv1567c"
/notes="Rv1567c, (MTCY336.36), len: 94. Unknown. Tbpase
score is 0.915."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1567c"
/protein_id="CAB09070.1"
/db_xref="GI:2117270"
/db_xref="SPTREMBL:O06623"
/translacion="NVTMTSMPSLRFAFTDNVCPDAPLVPFGVNYIYIPVMGGIG
AAIATVAGTGFVPSMLKGMYSFWPIISIAVITVTSIAGYAIAGFSERYWH"
3791..5104
/genes="bioA"
3791..5104
/genes="bioA"
/notes="Rv1568, (MTCY336.35c), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxononanoate
aminotransferase (EC2.6.1.62). O06622. Contains
aminotransferases class-III pyridoxal-phosphate attachment
site (P500600). FASTAresults, BioA.MYCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1%identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
identity in 421 aa overlap. Tbpase score is 0.874"
/codon_start=1
/transl_table=11
/product="bioA"
/protein_id="CAB09069.1"
/db_xref="GI:2117269"
/db_xref="SWISS-PROT:O06622"
/translacion="MAAATGGLTPEQIIIVDGAHIWIPIYSSIGREAVSPVVAVAIRGA
WLTIFRDQPIEVLDAMSWMWTAIHGHGHPALDQALPTQLKVMNVRWPGGITHGPAAR
LAKLIVDTTPAGLDTVFSDGSYSVEVAAKMALQYWKRLPKCKRRIWTWRGGYHGD
TFLAMSICDPHGHMSLWTDVLAQVFAPOVPRDYDPAYSAAFEAOLAOHAGELAAVV
VEPVQAGGRRHFRDPYRLHDLRIDICRYEVLIIFFDELATGHGRTGALFAADHAGVSP
DIMCVKALTGGYLSLAATLCTADVAHTISAGAAGALMGHPTFWANPLACAVSVASVE
LLGGDWRTITELAAGLTAGLDRALPAVTDVRVCGATGVIECDRPVDLAVATPAA
LDRGVLRFPFRLVYAMPPICTPAETITQITSAMVEVARLVGSLP"
4541..4654
/genes="bioA"
/notes="PS00600 Aminotransferases class-III
pyridoxal-phosphate attachment site"
5101..6261
/genes="bioF"
5101..6261
/genes="bioF"
/notes="Rv1569, (MTCY336.34c), bioF, len: 386. Function:
probable 8-amino-7-oxononanoate synthase (EC 2.3.1.47)
O06621. Contains aminotransferases class-II
pyridoxal-phosphate attachment site (PS00599). FASTA
results, BioF.MYCLE P45487(385 aa) opt: 1971; E(): 0;
80.1% identity in 381 aa overlap. Also similar to
MTCY1084.32, FASTA score: E(): 5.5e-29; 37.4% identity in
393 aa overlap"

```

```

/codon_start=1
/transl_table=11
/product="bioF"
/protein_id="CAB09068.1"
/db_xref="GI:2117268"
/db_xref="SWISS-PROT:O06621"
/translacion="MKAATQARIDISPLAWLDVAVQORHEAGRLRCLRPRAVATELD
LASNDYLGLSRIIPAVIDGGVQALRIWAGAGTSGRLVTDGTLHQOFFEALAEFFVGAAG
GILFSSGYTANLIGAVVGLSGPGLIIVSDARSHASLVDACRISHARVVVTPPIHVDVAD
AALRSKQDRAVVVTVDSVFSDGSLAPVRELEVCRRHGALLVDEAHGILGVKGGRG
LIYELCLGAPUVVMTTTLSSALSGOGVVLGPTPVRAHLIDAAKPP1FD7ICLAPAAV
GAARAAALVLOAEPMPQAVLNHAGELARMCGVAAVPSAMVSVILGEPESAVAAAA
CLDAGVKCCFEPPTVPAGTSRLRLUTARASLNAGELELARRVLTDVLAAVR"
5806..5835
/genes="bioF"
/notes="PS00599 Aminotransferases class-II
pyridoxal-phosphate attachment site"
6258..6938
/genes="bioB"
6258..6938
misc_feature
Query Match 90.2%; Score 408.4; DB 1; Length 32437;
Best Local Similarity 97.8%; Pred. No. 2.3e-56;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;
QY 1 galcggcgagggqacacatacaaaatgallcgcgcccltttccacacctgccgcgcggcg 60
|||||
DB 17861 GATCGCGGAGGGCGACGTCAAAGTGTATTCGTCGCCCTTTTTCGCCACCTGCCGCCGCGG 17802
QY 61 tggatgtgtlccarccgcagggcgccgaagcgacctgccgcgaagcg-ctcaatatac 118
|||||
DB 17801 TGGATGTGTCCACCGCCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17742
QY 119 gtccgcagagctgtgccgcgtacgccccagcggggtcatgactgggtacaccccgcgcg 178
|||||
DB 17741 GTCCCGAGCGAGTGTGCCCGCTACGCCACGCGGCTCATGACTGGCTACACCCCGACGCGG 17682
QY 179 acctcaccgacacgaaacgctgcgcgaacacgc-gcatcaccctgagcaccagcaatacg 237
|||||
DB 17681 ACCTCACCACACACGAACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17622
QY 238 atgcacatgtcacgctgactgactgactgactgactgactgactgactgactgactgactgact 297
|||||
DB 17621 AGCGCATGTTCATAGCTTAAGTGAATCATATCATATCATATCATATCATATCATATCATAT 17563
QY 298 gtgctagcacaactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 357
|||||
DB 17562 GTGCTAGCGCAACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17503
QY 458 gacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 417
DB 17502 GACACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 17443
QY 418 catgacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 453
|||||
DB 17442 CACGACGGGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17407
RESULT 10
MTCY251/c
LOCUS 38380 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 5/162.
ACCESSION Z74410 AL123456
VERSION Z74410.1 GI:3261600
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 38380)
REFERENCE
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,

```



```

78. .1481
/gene="Rv3448"
/notes="Rv3448, (MTCY77.20), len: 467 aa. Unknown membrane
protein, contains PS00402 Binding-comp signature. Some
transport systems inner membrane comp signature. Some
similarity to A021930|MTV035_18 Mycobacterium
tuberculosis (472 aa) fasta score, opt: 429 z-score:
488.4 E(): 1.1e-19; 28.2%identity in 479 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3448"
/protein_id="CAB08736.1"
/db_xref="GI:3261767"
/db_xref="SPTREMBL:O33354"
/translation="MPTSDGLRRVTHAGAAVDLTPAAVPVATLPSITVDLGR
CASPTARVOLSALGAPALPNATLTACQIRGQAVLVLHKSSAQPTPRCDQVAEAV
AAADTTARFQCRTRLSGALAASCITAGGLMLVRLNAGTNTVTRVSDATAGWAAA
TSVLAALFACGRTYRDIAGLTVLSTIFGAVAGLLVPGVPGVHSLVLAAMAAA
GLAMVITGSGTITFVACCAVVAATLVGAATAPVPAIGSLATLASFGLLVEA
ARMAVLGSLRPLPPALPDADALPTDRLTTRANKRADAWLTSLAAFAAATIGA
IGTAVATHGTHRSMSGGIALAAVTGALLLRARSADTRRSLVFAICITTTVAATFVA
ADRALEHPNIAALTAALAAVAFGLFVAPALSPTVYRTIELLECLALIAMVPLTA
WLCGAYSAVRHLDLTWT"
misc_feature
611. .>43401
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/db_clone="Y13E12"
1478. .2845
/gene="Rv3449"
1478. .2845
/gene="Rv3449"
/notes="Rv3449, (MTCY13E12.02), len: 455. Function:
probable precursor of serine protease. Has putative signal
peptideat N-terminus and hydrophobic stretch at
C-terminus. Contains three signatures typical of subtilase
family: aspartic acid active site (PS00136). histidine
active site (PS00137), serine active site (PS00138). FASTA
results: 053863 SERINEPROTEASE (390 aa) opt: 241; E():
8e-07; (38.0% identity in 387 aa overlap), similar to
MTCY15F10.29 (45.7% identity in 451 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3449"
/protein_id="CAB08735.1"
/db_xref="GI:2104372"
/db_xref="SPTREMBL:O06316"
/translation="MTTSRTLRLVLSALATLSGLTQVAHAVSPVPIDRWLPSEL
PAPRPVQREVCTEVAESGRFGRASQAQLADLDQVRLTRGACQVAV IDTQVA
RHRRLPKVAGDVFYTGCTGCTADCAHGTQVAGLIAAPUAGLNFSCVAPDVTLISI
ROSSKRPVDPSSDVGVDVTHAKAVKTAADLIGASVINISSIACVPAAPADRAL
GAALAYADVKNNAVIVAAAGTGAACPPQAPGVTKDSVTVAVSPAWYDDYVLTVGS
VNAQGPFAFTLAGPVDVAATGEATSLSPFGDGTVNRLLGQHGSIPIGTSYAAPV
VSGLAALIRARFPLTARQVMORIESTAHHPAGMDPLVNGTVDALAAVSSDSIPQA
GTATSDPAPVAVPVRSTPGPSDRRALHTAFAGAAICLLALMATLATASRRRLPRGN
GIAGD"
misc_feature
1757. .1789
/gene="Rv3449"
/notes="PS00136 Serine proteases, subtilase family,
aspartic acid active site"
1862. .1894
/gene="Rv3449"
/notes="PS00137 Serine proteases, subtilase family,
histidine active site"
2456. .2488
/gene="Rv3449"
/notes="PS00138 Serine proteases, subtilase family, serine
active site"
complement(2810. .4222)
/gene="Rv3450c"
complement(2810. .4222)
/gene="Rv3450c"
/notes="Rv3450c, (MTCY13E12.03c), len: 470. Function:

```

unknown but similar to Rv3895c MTCY15F10.17 (FASTA score: 1.6e-27; 36.4% identity in 475 aa overlap); possible membrane spanning region near N-terminus, and to Y14967|MLCB628_16 Mycobacterium leprae cosmid H628; (481 aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 32.9% identity in 480 aa overlap, also similar to Rv3869 (MTV027.04)"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3450c"
/protein_id="CAB08716.1"
/db_xref="GI:2104373"
/db_xref="SPTREMBL:O06317"
/translation="MPSPTATLVHSGYFLRRRIECALLFGDVCATGALRARTTSL
ALGCVLAIVAMGCAEVAILRPOSALGOAPLVMGRESGALYVVDVWHPVLNLASAR
LIAATNANPOPVSESELGHTKRGPLLIGCAPOLLDOPLAGAESAWAICDSNCGSTT
VVGPAEDSSAQNVTARQMLLVATESGSPTYLLXGGRRVVDLADPAVVALRLOGRV
PHVVGALLNAVPEAPRITAPRIGGRASVGLPFLVGGVVRITRASGEYVYVLED
QVOTIGQVADLLRFGDSQGVNVPTVAPDVRVAPLIVNTLPVSAPFDRPPTVDDGP
GRAVTLICVTTTAPQARVAFVLAUSGPVPLGGVPTLQAQDGRGPDALDADVLPFG
RSAYVAARSLSGGTGTGYLVTDITGVRFAIHDDDVADHLGLTAAIPAPVPVLATLPS
GPELSRANASVARIVTAPGB"
complement(4049. .4081)
/gene="Rv3450c"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(4228. .4231)
/notes="possible RBS, GGAG, for Rv3450c"
4388. .5131
/gene="Rv3451"
4388. .5131
/gene="Rv3451"
/notes="Rv3451, (MTCY13E12.04), len: 247. Function:
probable cutinase, similar to several, contains cutinase,
serine active site motif (PS00155). Alternative start
possible at 3733. FASTA results: CUT2_MYCTU_Q50664
probable cutinase GY339.08c precursor (219 aa) opt: 565;
E(): 2.3e-26; (44.8% identity in 223 aa overlap). Also
similar to MTCY13E12.05 (FASTA score: E(): 0; (59.2%
identity in 211 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3451"
/protein_id="CAB08717.1"
/db_xref="GI:2104374"
/db_xref="SWISS-PROT:O06318"
/translation="MAZALITAVVLLIATGAVTPVAFAGCPDAEFTVFARGTGEPP
GLIKVGGAPVNSLKKQGTMEIGVYVNVAAASRLQHGDDGANDAISHIKSMAGSCPTT
KLVAGYSQCATVITLIVAGVPLGSLIFGSLPLPAAYADNVAAVAFGNPSKAGSLSS
LSPLGSKALIDLLNFTLPIICIVGPGNEFSCHIDGYIPTTTQAAAFVQVRLRAGSVPI
LHNSVPIHGSVIGMPTAAVAPESLHGR"
misc_feature
4706. .4723
/gene="Rv3451"
/notes="PS00155 Cutinase, serine active site, GGYSGQ"
5165. .5168
/notes="possible RBS, GAGG, for Rv3452"
5178. .5858
/gene="Rv3452"
5178. .5858
/gene="Rv3452"
/notes="Rv3452, (MTCY13E12.05), len: 226. Function:
probable cutinase precursor. Contains PS00155 Cutinase,

```

Query Match 90.2%; Score 408.4; DB 1; Length 43401;
Best Local Similarity 97.8%; Pred. No. 2.1e-56;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

Qy 1 gatcgcgagcgccacatcaagtgattcgcccttttgcgcacacctgccgcgcg 60
|||||
Db 15625 GATCGCGAGCGCAGCTCAAGTGATTCGCGCCCTTTTTCGCCACCTGCCCGCGG 15684
|||||

Qy 61 tggatgtctccaccgccgagcgccgagcgacctgccgcaaaagc--ctcaatatc 118
|||||

Db	15685	TGGATGTGTCACCCGCCAGCGCGGCGGAAGCCGACGCTGGCCGGCAAGCCCGCTCAATATC	15744
QY	119	gtccgacgagctggcccgctacgcccagcgggctcatgactgggtacacccccgacggcg	178
Db	15745	GTCCGAGCAGCTGGCCCGCTACGCCACGCGGTGCTATGACTGGCTACACCCCGACGGCG	15804
QY	179	acctaccgacacccgacgcccgcgcaaacgc-gcatcacctgagcaacacgaataacg	237
Db	15805	ACCTACCGACACCGAACCGCCGCGGCAACCGCGGCATCACCTGAGCAACCGCAATACG	15864
QY	238	acggcatgtcacggctaaagtggctactgaccccccaagtgcgggcccacttttgaagcc	297
Db	15865	ACGGCATGTACCGCTAAGTGCTACTGACCCCCCAAG-CGGGGCCACCTTTGAAGCC	15923
QY	298	gtgtagcacaactggccgccccggcgcgacaaacccccgcagacaccccccggtcatc	357
Db	15924	GTGTAGCCAAACTGGCGCGCGCGGCGAGCAACCCCGAGCACACCCCGGTCAATC	15983
QY	358	gacacacccccgatggcgccgcctatcgaccgcgacacccgcagccaaagcccaacgcaac	417
Db	15984	GACACACCCCCGATGGCGCGGCCTATCGACCGCGCACCCCGAGCAACCCCAACGCAAC	16043
QY	418	cacgacgggctgctggccgggctgcgcgctgac	453
Db	16044	CACGACGGGTGCTGGCGGGCTGCG-TCGCTGATC	16079
RESULT	12		
AE007160			
LOCUS	16384 bp	DNA	BCT 27-APR-2001
DEFINITION	Mycobacterium tuberculosis CDC1551, section 246 of 280 of the complete genome.		
ACCESSION	AE007160	AE000515	
VERSION	AE007160.1	GI:13883389	
KEYWORDS			
SOURCE			
ORGANISM	Mycobacterium tuberculosis CDC1551. Mycobacterium tuberculosis CDC1551 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1. (bases 1 to 16384) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W. Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains Unpublished 2 (bases 1 to 16384) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W. Direct Submission Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. .16384 /organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" 96. .839 /gene="MT3557" 96. .839 /gene="MT3557" /note="similar to SP:Q00298; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="serine esterase, cutinase family"		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES	source		
gene			
CDS			

gene		/protein_id="AAK47897.1" /db_xref="GI:13883390"
CDS		/translation="MGAGALITAVVLLIALGAVWTLVAFADGCPDAEVTHARGTGEPP KIGRVQAFVDSLRQQTMEIGVYPVNYAASRLQDLHGGDGANDAI SHIKSMASCCPNT KIVLGGQSQAGTIDIVAGVPLGSI SFGSPLPAAYADNVAAVFQNPNSKAGCSLSS LSPUGSKAIDCNPTDPI CHVGPNEFSGHIDDI PTPTTQAASFVVVRLRAAGSVPH LPGSVPLPGSVLQMPCTAAPAPESLIHR"
	871. .1566	
	/gene="MT3559"	
	871. .1566	
	/note="MT3559"	
	/note="similar to SP:Q00298; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="serine esterase, cutinase family"	
	/protein_id="AAK47898.1"	
	/db_xref="GI:13883391"	
	/translation="MRPIGTIPRQPHSGRWAGARRLTSLVAAAFAAALLTLTPAL APASAGCPDAEVVFAFGTGEPPGLGRVGOAFVSSLRQNTKS IGTVGVNYPANGDFL AAADGANDASDHIOOMASACRATRLVGGY SOGAAVTDIVTAAPLPGLGTQPLPPAA IDHIAALALFCNPSGRAGGLMSALT PQGSKTINLCNNGDP ICSDGNRWRAHLGYVPG MTNQANFVASHI"	
	1624. .1863	
	/gene="MT3560"	
	1624. .1863	
	/gene="MT3560"	
	/note="identified by Glimmer2; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAK47899.1"	
	/db_xref="GI:13883392"	
	/translation="MTAADMHTAYHGLMLAPDPQAARRPSTOPNAHLSAKCAAMNRN HFTCKEVCGHANRTLGIT/TSVAIALKRYVDACHRNQ"	
	1845. .3523	
	/gene="MT3561"	
	1835. .3523	
	/gene="MT3561"	
	/note="identified by Glimmer2; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAK47900.1"	
	/db_xref="GI:13883393"	
	/translation="MMPGVITNSESPTAADHNRITATRETLEDYTLRLAPSKYRRWPP AVVGISAIGGIAYLAIDFAICANVGITWGTANALCGI AIFALVFEVTGLIPLYAARYN IDHDLITRGSQGYGYSWVTVIFATFTFIFFALESGSI MAQGLKGLHPI LMWAGYAS ILITPLVAVYGMKVLSQLIWTPLWLTLMAPFGYLAVSHIPSIGUFPFVAGDGHG GIPGVSLLAAWVLELLIAGLAQIIDLKPPHPTDENANHWPTLLAGCGWVAFQA LNGITGLPVLVLANIHGSSITLADQVPHQPMQIYRTFVPCWALATLAVILVLSQIK INVTNAYSLSLAWINSFTRPKHYGKRVVFGVNLAI AII MEANMFDLNTIIGCTA NTRMAWVAVASDUGPNKYILGLSFKTPEPRGMLYAINPVGFGSLIIAAGISIVTFP CGIAGAAQPPSPVAVIVTALVMPPI LAAATKGYYLRRTHDGLPLMYDEHGNLSAAV LITVCHQHFDPMLACOTHGAHVC SLCLSTDKOAEHLVPLGLARAHIPGIQVVP"	
	complement(3485. .4378)	
	/gene="MT3562"	
	complement(3485. .4378)	
	/gene="MT3562"	
	/note="similar to SP:P07649 GB:X02743 PID:41718 GB:U00096 PID:L788657; identified by sequence similarity; putative"	
	/codon_start=1	
	/transl_table=11	
	/product="urac pseudouridine synthase A"	
	/protein_id="AAK47901.1"	
	/db_xref="GI:13883394"	
	/translation="MSLTRPPPKSPQRRPRISGVVRLRLDIADYDGTDFAGWAAQVQGV RTVAGDLDAALTTFRTDVRRLRAGRTDAGVHASQVAHVDPADALPNAYPRAGHVG DPEFLPLRLRLRGLPADVRLDITRAPAGDFARFSALRRHYVRLSTPYGVPEPOQA KYITAMPRELDUDAMTAASRLMGLHDFAEFCRHREGATTIRDLQRIDMSRAGTLVTA HYTADAFCSWRSRLVGLLAVGEHRRATTWCRELTLATGRSSDFAVAPAIGLITLIQV DYPDDQLASRLNLTVDVRSG"	
	complement(4323. .4865)	
gene		

```
/gene="MT3563"
complement(4323. .4865)
/gene="MT3563"
/notes="similar to GB:J01685 SP:P02416 GB:X00766 PID:147716
PID:42800; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L17"
/protein_id="AAK47902.1"
/db_xref="GI:13883395"
/translation="MPKPTKGRLLGGSSSHQKAILANLATSIFEHGRITITTEPKARAL
RPAKGLTHAKKALNRRREVLKLRDKDVVHTLFAEIGPFFADRDGGYTRIKIEA
RKGDNAPNAVTELVEKTEVTSEANRRARVAAQAQAKAKAAAMPTESEAKPAEEGDVV
CASEPDAAKPEPTEAPEN"
complement(4897. .5940)
/gene="MT3564"
complement(4897. .5940)
/gene="MT3564"
/notes="similar to SP:P72404; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-directed RNA polymerase, alpha subunit"
/protein_id="AAK47903.1"
/db_xref="GI:13883396"
/translation="MLISQRPSTLSDVLTNRQSVIEPIEPGFYTLIGNSLRRTLLS
SIPGAATVSIIRIDGLVHEFTVPGVKEDVTEIILNLSLVVSSSEDEPVTWYLRKQGP
GEVAGDILVPAGTVVHNPQMHIATLNDKGLVELVVERCKGYVPAVONRASGAETG
RIPVDSIYSPVLKVTYKVDATRVQRTDFDKLILDVETKNSISPRDAISAGRTLVEL
FGLAEHLNVEAGEIEIGSPAEDHIAFPIDDLDTVRSYNCLKREGVHTVGELY
ARTESDLIDIRNFQGSIDEVKIKLHQLGLSLKDSPFSPFSEVAGYDVATGTWSTEG
AYDEQVTAETQL"
complement(6092. .6697)
/gene="MT3565"
complement(6092. .6697)
/gene="MT3565"
/notes="similar to SP:P81288; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAK47904.1"
/db_xref="GI:13883397"
/translation="MARYTGPVTRKSRRLRTDLVGGDAFEKRPYPQGHGRARIKES
EYLLOQEKQARFTYGVMEKOFRTYVEAVRQPKTGEELKILLESKLDNVIYRAGL
ARTREMAROLVSHGHNNGVHVNVPVSRVSOYDIDVDRDKSLATVPQIARETAGER
PIPSWLQVWGERQVLIHQLPERAQIDVPLTEQLIVEYSK"
complement(6706. .7125)
/gene="MT3566"
complement(6706. .7125)
/gene="MT3566"
/notes="similar to SP:P45812; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S11"
/protein_id="AAK47905.1"
/db_xref="GI:13883398"
/translation="MPPAKKGPATSAKRGOKTRRREKKKNVPHGAHIKSTFNNTIVTI
TDPGQNVIANASSGHVGFKSRKSTPFAAQLAENAAKKAQDHGVRKVDVFKGPGSG
RETAIRSLQAAGLVEGALSDVTPQPHNGVRPPKRRRV"
complement(7129. .7503)
/gene="MT3567"
complement(7129. .7503)
/gene="MT3567"
/notes="similar to GP:2104383; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S13"
/protein_id="AAK47906.1"
/db_xref="GI:13883399"
/translation="MARLVGVDLPRDKRMEVALTYIFGIGRTSRSEILATGIDRDLR
```

```
TRDTEEQIHLRIDYIEANLKVEGDLRREVQADIRRKIEIGCYOGLHRRGMPVRQR
TKTNARTKGPRTIAGKKAR"
complement(7719. .7832)
/gene="MT3567.1"
complement(7719. .7832)
/gene="MT3567.1"
/notes="similar to SP:P45810; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11

Query Match      89.1%; Score 403.6; DB 1; Length 16384;
Best Local Similarity 97.1%; Pred. No. 1.8e-55;
Matches 443; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

QY 1 gatcgccgaggtccacatcaaatatgatctcgcctcttttgcacacactgcgcgcgcg 60
|||||
DB 11332 GATCGCGGAGGCGCGCTCAAGTCAATTCGCGCCCTTTTTCGCCACCTGCCCGCGCGG 11391
QY 61 tgaatgtatcttcacgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 118
|||||
DB 11392 TCGATGTCTCCACCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAATATC 11451
QY 119 gtccgagaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 178
|||||
DB 11452 GTCCCGACGAGTTCGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11511
QY 179 aactcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 237
|||||
DB 11512 ACCTCACCAGACACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11571
QY 238 acggcatgtcacgcgttaagtcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 297
|||||
DB 11572 ACGGTATGTACCGCTAAGTGGCTACCTGACCCCGCGCGCGCGCGCGCGCGCGCGCG 11630
QY 298 gtgctagcacaactggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 357
DB 11631 GTGCTAGCGCAACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11690
QY 358 gacacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 417
DB 11691 GACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11750
QY 418 cagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 453
|||||
DB 11751 CACGACGGGCTCTCTGGCGCGCGCTGCACCGCGCGCGCGCGCGCGCGCGCGCGCG 11786

RESULT 13
LOCUS 186264 12412 bp DNA PAT 10-JUN-1998
DEFINITION Sequence 18 from patient US 9/00683.
ACCESSION 186264
VERSION 186264.1 GI:32059W2
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified
REFERENCE 1 (bases 1 to 12412)
AUTHORS Stover,C.Kendall and Mahairas,G.G.
TITLE Virulence-attenuating genetic deletions deleted from mycobacterium
BCG
JOURNAL Patent: US 5700683-A 18 23-DEC-1997;
FEATURES location/Qualifiers
source 1..12412
/organism="unknown"
BASE COUNT 2146 a 4274 c 3875 g 1917 t 200 others
ORIGIN

Query Match      81.0%; Score 366.8; DB 6; Length 12412;
Best Local Similarity 95.6%; Pred. No. 1.3e-49;
Matches 432; Conservative 0; Mismatches 12; Indels 8; Gaps 5;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 19:17:10 ; Search time 2977.34 seconds
(without alignments)
7718.491 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcgaa.....cgacgacaaacccgattgac 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_ov:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pl:**

25: em_ro:**

26: em_sts:**

27: em_sy:**

28: em_un:**

29: em_vi:**

30: em_htgo_hum:**

31: em_htgo_inv:**

32: em_htgo_rod:**

33: em_htgo_hum:**

34: em_htg_inv:**

35: em_htg_rod:**

36: em_htg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1393	100.0	43401	1	MTY13E12	295390 Mycobacteri
2	1354.2	97.2	3453	1	MTU143540	U43540 Mycobacteri
C 3	1320.6	94.8	17783	1	AF007028	AE007028 Mycobacte
C 4	1315.8	94.5	10019	1	AF041819	AF041819 Mycobacte
C 5	1315	94.4	16384	1	AE007160	AE007160 Mycobacte
C 6	1314.2	94.3	1604	1	MBU35021	U35021 Mycobacteri
C 7	1307.4	93.9	32437	1	MTY3336	295586 Mycobacteri
C 8	1301.4	93.4	38380	1	MTY251	274410 Mycobacteri
C 9	1298.2	93.2	5764	1	AE006921	AE006921 Mycobacte
C 10	1281.8	92.0	9281	1	MBU4351	U35017 Mycobacteri
C 11	1202.4	86.3	12412	6	186264	186264 Sequence 18
C 12	699.6	50.2	34331	1	MTCL165	295584 Mycobacteri
C 13	698	50.1	14953	1	AE006996	AE006996 Mycobacte
C 14	683.6	49.1	33100	1	MTY9F9	284498 Mycobacteri
C 15	682	49.0	14179	1	AE007053	AE007053 Mycobacte
C 16	675.6	48.5	40056	1	MSGY348	AD000702 Mycobacte
C 17	636.8	45.7	15019	1	AE007036	AE007036 Mycobacte
C 18	636.8	45.7	37432	1	MTCL125	298268 Mycobacteri
C 19	579.4	41.6	15504	1	AE006995	AE006995 Mycobacte
C 20	579.4	41.6	22550	1	MTY2208	295585 Mycobacteri
C 21	455	32.7	1535	6	AR096713	AR096713 Sequence
C 22	455	32.7	1535	6	AR147694	AR147694 Sequence
C 23	396	28.4	650	6	AR096715	AR096715 Sequence
C 24	396	28.4	650	6	AR147696	AR147696 Sequence
C 25	238	17.1	43254	1	MLCB1779	298271 Mycobacteri
C 26	238	17.1	312050	1	MLEPRTN3	AL583919 Mycobacte
C 27	212.8	15.3	348450	1	MLEPRTN1	AL583920 Mycobacte
C 28	137	9.8	36138	1	MLU15187	U15187 Mycobacteri
C 29	137	9.8	38109	1	MLCH5	295151 Mycobacteri
C 30	137	9.8	342300	1	MLEPRTN8	AL583924 Mycobacte
C 31	135.2	9.7	39193	1	U00021	U00021 Mycobacteri
C 32	135.2	9.7	332450	1	MLEPRTN5	296801 Mycobacteri
C 33	114.2	8.2	36225	1	MLCL581	AL583918 Mycobacte
C 34	114.2	8.2	344050	2	MLEPRTN2	Continuation (4 of
C 35	111.8	8.0	56870	2	CEY3986_3	Continuation (4 of
C 36	108.6	7.8	210827	2	AC090654	AC090654 Mus muscu
C 37	107.4	7.7	312050	2	MLEPRTN4	AL583919 Mycobacte
C 38	101.2	7.3	16387	1	AE006920	AE006920 Mycobacte
C 39	101.2	7.3	29256	1	MTV030	AL021428 Mycobacte
C 40	98.8	7.1	47852	2	MTV023	AL021428 Mycobacte
C 41	97.8	7.0	303091	2	AC084799	AC084799 Mus muscu
C 42	96.8	6.9	152547	2	ALI37857	ALI37857 Homo sapi
C 43	96.8	6.9	177883	2	AC046159	AC046159 Homo sapi
C 44	96.2	6.9	158167	2	AC044105	AC044105 Homo sapi
C 45	94.6	6.8	14472	1	AE003928	AE003928 Xylella f

ALIGNMENTS

RESULT	1
LOCUS	MTY13E12
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 147/162.
ACCESSION	295390 AL123456
VERSION	295390.1 GI:3261766
KEYWORDS	
SOURCE	Mycobacterium tuberculosis H37Rv.
ORGANISM	Mycobacterium tuberculosis H37Rv. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex. 1 (bases 1 to 43401)
REFERENCE	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,

Squares, S., Sqaures, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
 Nature. 393 (6685), 537-544 (1998)
 98295987
 2 (bases 1 to 43401)
 Parkhill, J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2104370.

COMMENT

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of 178 genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

Location/Qualifiers
 1. 43401
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 <1. 1481
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="y77"
 78. 1481
 /gene="RV3448"
 78. 164
 /gene="RV3448"
 /note="PS00402 Binding-protein-dependent transport systems:inner membrane comp signature"
 78. 1481
 /gene="RV3448"
 /note="RV3448, (MTCY77.20), len: 467 aa. Unknown membrane protein, contains PS00402 Binding-protein-dependent transport systems:inner membrane comp signature. Some similarity to AL021930|MTV035_18 Mycobacterium tuberculosis (472 aa) fasta score, opt: 429 z-score: 488.4 E(): 1.1e-19; 28.2% identity in 479 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3448"
 /protein_id="CAB08736.1"
 /db_xref="GI:3261767"
 /db_xref="SPTREMBL:O33354"
 /translation="MPTSDGLRRVTVHAGAAQVDLTLPAAVPVATLIPSTVDILGDR GASPTARYOLSLGAPLPNATLAQCGRDCAVLVLHKSSAQPTPRCDDDVAEAV AALDTARPOQRTLRISGALAASCITAGGGLMLVRNALGTNTVTRYSDATAGVVAAG LALLFAVIACRTYRDP IAGLTLSVATIFGAVAGLLAVPGVGVHSLVAAAMAAA TSVLAMRTGGGTLTAVACCAVVAATLVGATTAAPVPAIGSLATLASFGLLEVS ARMYVLLAGLSPRLPALNPADADLTDRLTRANRADAWLTSLAAFAATGTA IGTAVATGTHRRSMGGIALAAVTGALLLRASADTRRSLVFAICITTVATATTA ADRALEHCPWTAALTAALAAVAMFLGFVAPALSLSPTVYRTIELLECLALIAWPLTA WLCGAYSAVRHLDLTWT"
 611. >43401
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"

gene

CDS

/db_xref="taxon:83332"
 /clone="Y13E12"
 1478. 2845
 /gene="RV3449"
 1478. 2845
 /gene="RV3449"
 /note="RV3449, (MTCY13E12.02), len: 455. Function: probable precursor of serine protease. Has putative signal peptide N-terminus and hydrophobic stretch at C-terminus. Contains three signatures typical of subtilase family: aspartic acid active site (PS00136), histidine active site (PS00137), serine active site (PS00138). FASTA results: Q53863 SERINEPROTEASE (390 aa) opt: 241; E(): 8e-07; (38.0% identity in 387 aa overlap), similar to MTCY15F10.29 (45.7% identity in 451 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3449"
 /protein_id="CAB08745.1"
 /db_xref="GI:2104372"
 /db_xref="SPTREMBL:O06316"
 /translation="MTTSKTLRLVVSALATLSGLTPVAHAVSPPIDERWLPESAL PAPRPTVOREVCTEVTAEGRAGRAESAOALADLDVMRLTRGAGRVAVIDTQVA RHRLPKVYAGDYVFTGDTACDCAHGLVAGIIAAPDAQSDNPSGVAPDVLTISI QSSSFAPVDPSSTVGVDVDTMAKAVRTAADLGASVINISSIACVPAARAPDDRAL QAAAYAVDKNAIVVAAGNTGGAQCPQAPGVTRDSVTVAVSPATDDUVTLYTGS VNAQGPSAFTLAGPMDVVAATGEAVTSLSPFGDGTNRLGGHGSIPISGTSYAAAP VSCIAALIRARFPTLIRQVMQRIESTAHHPAGMDPLVGNCTVDALAAVSSDSIPOA GIATSDPAPVAVPVPRSTPGPSDRRLHTAFAGAAICLLMATLATATASRLRPGRN
 1757. 1789
 /gene="RV3449"
 /note="PS00136 Serine proteases, subtilase family, aspartic acid active site"
 1862. 1894
 /gene="RV3449"
 /note="PS00137 Serine proteases, subtilase family, histidine active site"
 2456. 2488
 /gene="RV3449"
 /note="PS00138 Serine proteases, subtilase family, serine active site"
 /complement(2810. 4222)
 /gene="RV3450c"
 /complement(2810. 4222)
 /note="RV3450c, (MTCY13E12.03c), len: 470. Function: unknown but similar to RV3895c MTCY15F10.17 (FASTA score: 1.6e-2; 46.4% identity in 475 aa overlap); possible membrane spanning region near N-terminus, and to Y14967|MLJ0628 16 Mycobacterium leprae cosmid H628; (481 aa) opt: 708 z-score: 710.9 E(): 4.5e-32; 32.9% identity in 480 aa overlap, also similar to RV3869 (MTV027.04)"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv3450c"
 /protein_id="CAB08716.1"
 /db_xref="GI:2104373"
 /db_xref="SPTREMBL:O06317"
 /translation="MPSPTATLHVSGYRELLRRIECALLFGDVCATGALRATTSL ALGCLVAIVAMGCAVALLRPQALQAPIVMGRESGALYVRVDVWHVPLNLASAR LIAATNPAVPMSESELGHTKRGPLIGIPGAPOLLQDPLAGAESAWAICDSNCGSTT VVVGPAEDSSAQLVTAQOMILVATESGSPTYLLYGGRRVVDLADPAVVALRQGRV PHVVAQSLNVAPEAPRITAPRIKGRASVGLYGGVGVVIRITRASGDYVYVLSD GVQIGVQADLRLFGDSQGSQSNVPTVADPVRVAPVYNTLPSAFEDRPTVDGSP GRVATTLCTVTTTQAQCAARVAVLAGSPVPLGVGVVTLQAQDRGCPALDAVLYLPG RSYAVAAASLSGGGTGRYLYLTDVTGVRFAIHDDDDVAHDLGLTAAIPAPMPVLTLPSP GPELSRANASVARDTVAPGP"
 /complement(4049. 4081)
 /gene="RV3450c"
 /note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

misc_feature

misc_feature

misc_feature

gene

CDS

misc_feature

source

ORGANISM	Mycobacterium tuberculosis H37Rv
REFERENCE	1. (sites)
AUTHORS	Lee, T.
TITLE	Rhamnose biosynthetic genes related to a novel repeated sequence of Mycobacterium tuberculosis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3453)
AUTHORS	Lee, T.
TITLE	Direct Submission
JOURNAL	Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
REFERENCE	3 (bases 1 to 3453)
AUTHORS	Lee, T.
TITLE	Direct Submission
JOURNAL	Submitted (14-AUG-1997) Tae-Yoon Lee, Microbiology, College of Medicine, Yeungnam University, 317-1, Daemyungdong, Namku, Taegu 705-035, South Korea
REMARK	Sequence update by submitter
COMMENT	On Aug 14, 1997 this sequence version replaced gi:1304703.
FEATURES	Location/Qualifiers
source	1..3453 /organism="Mycobacterium tuberculosis" /db_xref="taxon:1773" /clone="pIS116" 150..1139 /gene="rfba" 150..1139 /gene="rmb" 150..1139 /gene="rmlb" /note="previously known as rfbA" /codon_start=1 /transl_table=11 /protein_id="AAB66648.1" /db_xref="GI:2326950" ADVEDAIRLVGGDITDAELVSQLVAESDAVVHFAESHVDNALDNPEFLHTNIGTF TILEAVRRHGRVLRHHISTDEVYGOLELDDRRFTSTPYNPSPYSATKACADMLVRA WVSRLCVRATISNCSNNYPYQHVKEFTPTNCLTGLKLYGDDGNVRDMIHVDH NSAVRILDRIGRIGETYLISSEGERDNLTVLSRLMLDRDPDFDHTVRAVDLRY AIDPSRLDELCAWPKHTDFERALTITIDWTRDNESLWLSUKHAGGRYQDAQ" 150..1139 /gene="rmlb" /codon_start=1 /transl_table=11 /product="rhamnose biosynthesis protein" /protein_id="AAB66647.1" /db_xref="GI:2326949" ADVEDAIRLVGGDITDAELVSQLVAESDAVVHFAESHVDNALDNPEFLHTNIGTF TILEAVRRHGRVLRHHISTDEVYGOLELDDRRFTSTPYNPSPYSATKACADMLVRA WVSRLCVRATISNCSNNYPYQHVKEFTPTNCLTGLKLYGDDGNVRDMIHVDH NSAVRILDRIGRIGETYLISSEGERDNLTVLSRLMLDRDPDFDHTVRAVDLRY AIDPSRLDELCAWPKHTDFERALTITIDWTRDNESLWLSUKHAGGRYQDAQ" 1136..1747 /gene="rmlc" 1136..1747 /gene="rmlc" /note="previously known as rfbC" /codon_start=1 /transl_table=11 /protein_id="AAB66649.1" /db_xref="GI:2326951" LDVRQVNCSSVAGVLRGLHFAQLPPSOAKYVTCVSGSVDFDVVDIREGSPTFRWDS VLLDDDRNIYSDGLAHGLFALQDNSTVMYLCSAEYNPQROHTICATDPTLASIRRW SDGGCPSLSDYDAAPSFEDVRRLGLLPGGTIVSRF1GEMRGT" 2202..2655
repeat_region	

BASE COUNT	679 a	1226 c	990 g	558 t
ORIGIN				
Query Match	97.2%;			
Best Local Similarity	99.2%;			
Matches 1382;	Conservative	0;	Mismatches	8;
			Indels	3;
			Gaps	2;
Qy	1	1	1	1
Db	1831	1	1	1
Qy	61	1	1	1
Db	1891	1	1	1
Qy	121	1	1	1
Db	1951	1	1	1
Qy	181	1	1	1
Db	2009	1	1	1
Qy	241	1	1	1
Db	2069	1	1	1
Qy	301	1	1	1
Db	2129	1	1	1
Qy	361	1	1	1
Db	2189	1	1	1
Qy	421	1	1	1
Db	2249	1	1	1
Qy	481	1	1	1
Db	2409	1	1	1
Qy	541	1	1	1
Db	2368	1	1	1
Qy	601	1	1	1
Db	2428	1	1	1
Qy	661	1	1	1
Db	2488	1	1	1
Qy	721	1	1	1
Db	2548	1	1	1
Qy	781	1	1	1
Db	2608	1	1	1
Qy	841	1	1	1

/note="3-4 copies in M. tuberculosis genome; Found in the rfb cluster (inside the rfbA gene), also found in the mce of M. tuberculosis; present at 3-4 sites of the genome of H37Rv, H37Ra, Erdman, and many isolates; not found in M. smegmatis & M. avium."
/rpt_type=dispersed
/rpt_unit=1806..2250

|||||
Db 6280 GTCATGACACACCCCGATGCGCGCATCGACCGCGACACCGCGACGCCAA 6221
Qy 787 cgaacacagaggggtgtgtgcccgggtgagcgagctgacgctcccggaactgggc 846
Db 6220 CGAACACAGAGGGGTGTGCGCGGGCTGCGCGCGCTGATCGCTCCCGGAACGTGGC 6161
Qy 847 caacacaaaggtcttccctctcgtatcgtgtggtcaccacacacccctgacgacactgcaaac 906
Db 6160 CAACACAGGGTCTCCCGTCTCGATCGTGGTACACACACCCCTGACCGACCTGCAACAC 6101
Qy 907 ggcgcgcgaaggggttaccggcgccgagcaccctgtaccatggccgagtgatccgc 966
Db 6100 GCGCGCGGAAGGGTTCACCGCGCGCGCACCTGCTACCCATGCGCGATGTATCCGC 6041
Qy 967 atgacagcagccacacactactctcccccgaagcgagagtagtaccacagcgatcttc 1026
Db 6040 ATGACGAGCGACGCCACCACTACTCCCGCGAAGCGGAGGTACCCGACGCGATCTTC 5981
Qy 1027 gaccagggcacacccctggcggtgtatcatcaccaaacgcctagcctccccggccagcgg 1086
Db 5980 GACCAGGACACCCCTGGCGGTGTATCACACCAACCGCTAGCCTCCCGGCCACGCGG 5921
Qy 1087 atcatgtcttgcgaagcagcgagctgcacaaacccggctgtgacgacccgacctac 1146
Db 5920 ATCATGCTGTCGCAACAGACCGCGCTGCACCAACCCGGCTGTGACGACCGCCCTAC 5861
Qy 1147 cacagcaagccacacagtcacccgctggacgacgacgagcagcgacgacatcacgag 1206
Db 5860 CACAGCAAGCCACACGTCACCGGCTGGACACGACCGGACGACCGACATCACCGAC 5801
Qy 1207 ctgacctgctgctggcgccgagacacgagactcgccgaaaaggctgacacccacac 1266
Db 5800 CTCACCTGGCTGGCGCCGACACACGACTCGCGAAGAGGCTGACACCGCCGCAAA 5741
Qy 1267 aacacccagcgccacacggaatggctacacacacccacccctgcacgagcgcaacccgc 1326
Db 5740 AACACCCAGGCGCACACGATGGTACCAACCCCGACCTCGACCGACCGCAACCCGCG 5681
Qy 1327 accaaccttcacaccccggaagctctctcccaacacgagcagcagcaaaccc 1386
Db 5680 ACCAACCTTCCACACGAGAGTGTGTACGCCCAACACGAGGAGGACCAACCGAC 5621
Qy 1387 gat 1389
|||
Db 5620 GAT 5618

RESULT 5
AE007160
LOCUS
DEFINITION
Mycobacterium tuberculosis CDC1551, section 246 of 280 of the complete genome.
ACCESSION
AE007160 AE000516
VERSION
AE007160.1 GI:13883389
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Unayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
Unpublished
2 (bases 1 to 16384)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,

Kolonay, J.F., Nelson, W.C., Unayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..16384
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:8331"
/note="clinical strain"
96..839
/gene="MT3557"
96..839
/gene="MT3557"
/note="similar to SP:Q00298; identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47897.1"
/db_xref="GI:13883390"
/translation="MGAGALITAVVLLIALGAVTVLVAFDGCPDAEVTFARGTGEPPGICRGVGFVDSLRQQTGMEIGVYVPAASRLQHGCDGANDAISHIKSMAGSCPNTKLVLGGYSOGATVIDIVAGVPLGSIISFGSLPAAYADNVAVAFGPNRAGGSLSSLSPLFGSKAIDICNPDPICHVGPNGNEFSGHIDYIFPTTQAASFVVQIKAGSVPHLPQSVPLPGSVLQMPCTAAPESLHGR"
871..1566
/gene="MT3559"
871..1566
/gene="MT3559"
/note="similar to SP:Q00298; identified by sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="serine esterase, cutinase family"
/protein_id="AAK47898.1"
/db_xref="GI:13883391"
/translation="MRFGIVIPRQPHSGRWRAGAAARLTSLVAAAFATLLTPALAPASAGCPDAEVVFARGTCEPFGICRGVGFVDSLRQQTGMEIGVYVPAASRLQHGCDGANDAISHIKSMAGSCPNTKLVLGGYSOGATVIDIVAGVPLGSIISFGSLPAAYADNVAVAFGPNRAGGSLSSLSPLFGSKAIDICNPDPICHVGPNGNEFSGHIDYIFPTTQAASFVVQIKAGSVPHLPQSVPLPGSVLQMPCTAAPESLHGR"
1624..1863
/gene="MT3560"
1624..1863
/gene="MT3560"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47899.1"
/db_xref="GI:13883392"
/translation="MTAADWHTAYHGLMAPDPAQARRPSTQPNNAHLSAECAAMRNHFTGREVCGHANRTLTGTATATLRVDAGRHKKQ"
1835..3523
/gene="MT3561"
1835..3523
/gene="MT3561"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47900.1"
/db_xref="GI:13883393"
/translation="MMPGVITNSPSTADHDRTATRTLEDYTLRLAPRSYRRWPPAVVGSALGCIAYLADFAFANVTGWGTANALCGIAPLVVFTGLPAYAARYNIDLDLITSGSGFYGVSVTVNIFATFTFFALEGSIQAQGLKGLHILWAGYACS TLIIFPLVYVYGMKLSQLQMTPLWILMAAPFGVLYVSHDPSTGQFYSYAGKDHG GLSGSVLLAAGVCLSLTAQIAEQIDYLRFPMPRTPENANRWMTWTLIAGPCWAFGA TKOITGLFLAVYLMANIPGSSSTIANOPVHQMIVTETVPGHIALTLAVILVLSQIK INVTNAYSGLAWTNSFTRLTKHYFGRVVFLGVNLAIILMEANFDLNTILGCTA

[illegible]

[illegible]

Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, Meudon-la-Forêt Cedex 15, France E-mail: parkhill@sanger.ac.uk
75724-937, 1998; this sequence version replaced 01:2117233.

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: in some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

```

/ gene="Rv1566c"
/ note="Rv1566c", (MTCY336.37), len: 230. Function: unknown
probably exported has QQAPV repeats at C-terminus, similar
to hypothetical M. leprae protein, to putative invasin
1.2(007390, 007391) from M. avium and slightly similar to
C-terminus of Listeria invasion-associated protein p60
precursor p60.LISMO P21171. FASTA results, Q49634 COSMID
B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
overlap. TBparse score is 0.890"
/ codon_start=1
/ trans1_table=11
/ product="hypothetical protein Rv1566c"
/ protein_id="CAB09071.1"
/ db_xref="GI:2117271"
/ db_xref="SPTREMBL:O06624"
/ translation="MKRSKSGSFAIGLAMMLAPMWAAAGLAAADPATRPVDYQQIITD
VVIAGLSORGVPFSGAGGIGSGPTRGTTGINTVGFDSAGLIQYAVAGAGLKLPRSS
GOMYKQGVQLPQARKGDLIFYGPEGTQSVLYLGKQMLEVDGVQVQSPVRNGMT
PYLVLTGQPTPVQQAAPVQAPVQQAAPVQQAAPVQQAAPVQQAAPVQQAAP
VQPPPGTAGSR"
complement(3030, 3036)
/ note="possible RBS, AAGAGG, for Rv1566c"
complement(3258, .3263)
/ note="ASNI site: ATTAAT; probably linking fragments
B3/66"
complement(3259, .3543)
/ gene="Rv1567c"
complement(3259, .3543)
/ gene="Rv1567c"
/ note="Rv1567c", (MTCY336.36), len: 94. Unknown. TBparse
score is 0.915"
/ codon_start=1
/ trans1_table=11
/ product="hypothetical protein Rv1567c"
/ protein_id="CAB09070.1"
/ db_xref="GI:2117270"
/ db_xref="SPTREMBL:O06623"
/ translation="MVTMTSNPSRLFAFTNVCPDPLVPFGVNYIYIPVMWGGIG
AAIATAVIGPFSMLKGMWMSFWPIISIAVITVTSIAGYAIAGFSRYNH"
3791. .5104
/ gene="bioA"
3791. .5104
/ gene="bioA"
/ note="Rv1568", (MTCY336.35c), bioA, len: 437. Function:
probable adenosylmethionine-8-amino-7-oxonanoate
aminotransferase (EC2.6.1.62) O06622. Contains
aminotransferases Class-III pyridoxal-phosphate attachment
site (PS00600). FASTA results, BIOA_MyCLE P4548 (436 aa)
opt: 2534; E(): 0; 85.1% identity in 436 aa overlap. Also
similar to other M. tuberculosis proteins e.g.
MTCY227.12c, (449 aa). FASTA score: E(): 3.5e-16; 29.5%
identity in 421 aa overlap. TBparse score is 0.874"
/ codon_start=1
/ trans1_table=11
/ product="bioA"
/ protein_id="CAB09069.1"
/ db_xref="GI:2117269"
/ db_xref="SWISS-PROT:O06622"
/ translation="MAAATCGLTPEQIIAIVDGAHLWHPSYISIGREAVSPVVAAHGA
WFLIIRDGPRIEVLDAWSSMTAIHGHGHPALDQALTOLRYNMHVMFGGLTHEPAAR
LAKLLVDITPAGLDVFFSDSGSVSVEVAAKKALQYWRGRLPGKRLMTWRRGGYHGD
TFPLMSICDPHGSMHSLTWDVLAQVAFQVPRDYPAYSAFAQLQWAGELAAVY
VEPVQAGAGMRHDPRLYLHLDRICRRYEVLLIIEIATGFGRTGALFAADHAGVSP
DINCVQAGTGLTSLAATLCTADVAITLSAGAALMHGPTFMANPLACAVSVASVE
LLIGDWMRTRI TELAAGLAPGLADPAVTDVVRVCGAIGVIECDRPVDLAVATPAA
LDGWVLRFLRNLVAMPPIYICTPAEITQITSAMVEVARLVGSLP"
4541. .4654
/ gene="bioA"
/ note="PS00600.0 Aminotransferases class-III
pyridoxal-phosphate attachment site"
5101. .6261
/ gene="biof"
5101. .6261
CDS

```

```

/ gene="bioF"
/ note="Rv1569", (MTCY336.34c), bioF, len: 386. Function:
probable 8-amino-7-oxonanoate synthase (EC 2.3.1.47)
O06621. Contains aminotransferases class-II
pyridoxal-phosphate attachment site (PS00599). FASTA
results, BIOF_MyCLE P45487(385 aa) opt: 1971; E(): 0;
80.1% identity in 381 aa overlap. Also similar to
MTCY10H4.32, FASTA score: E(): 5.5e-29; 37.4% identity in
393 aa overlap"
/ codon_start=1
/ trans1_table=11
/ product="bioF"
/ protein_id="CAB09068.1"
/ db_xref="GI:2117258"
/ db_xref="SWISS-PROT:O06621"
/ translation="MKAATQARIDDSPLAWLDVQVQRHEAGLRCLRRPAPAVATELD
IASNDVYLGLSRHPFAVIDGGVQALRIMGAGATGSRVLTGDKTLHQFEAEAEFVGAA
GLLFSSGYTANLGAIVGAVSLGSGPGLLVSDASLASLVDRCLRSRVVVTGPHRDVDAVD
AALRSDEQRADVVTDSVFSADGSLAPVRELLVCRHGLALLDEAHIGLVGGRRG
LLYELGAGAPDVVMTTLLSKGSLGSGGVVVGPTVRAHLIDAARPIFDTGLAPAAV
GAARAALRVIQAEPMRPQAVLNHAGFLARMCGVAAPDSAMYSVLLGPSPESAFAAAA
CUDAGVKVGCFRPPTVPAGTSLRLTLARASLNAGLELARRVLTDLVLAARR"
5806. .5835
/ gene="bioF"
/ note="PS00599.0 Aminotransferases class-II
pyridoxal-phosphate attachment site"
6258. .6938
/ gene="bioD"
6258. .6938
misc_feature
Query Match 93.9%; Score 1307.4; DB 1; Length 32437;
Best Local Similarity 98.4%; Pred. No. 2.5e-169;
Matches 1320; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 7 cgggtagccgcgaacgaggttcgaggtcttttgatcgctggatgcgagctgagcgct 66
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18228 CGAATAGCCGGGAGGACCTGTGCGAGCTTTTGATCGCTGGATGCGAGCTGACCGCT 18169
QY 67 tggacgaggtgctcttttgaggtgttggaccaccccaagacggctgcggtctctggaacgctc 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18168 TGGACGAGGTGTCTTTTGAGGTGTTGACCACCCCGGAGCGGTCTCTTGAACGCTC 18109
QY 127 tggaaatgcttggcgcgcgcgtacccggttgggtcacgctgtatcaaccacacttgacg 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18108 TGGAAATGCTTGGTCGCCCGGCTACCCGCGGTGGGTCAACCGTTGATCAACCACTTGACG 18049
QY 187 cccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18048 CTTCAAGCTAGCTAGTAAAGAAATTTGAAATGATCTGCTGCGCGCTGGCCACACCGGTTAC 17989
QY 247 gcatcaatcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17988 GCATCACCAGCCCGGACCGCCCGCGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17929
QY 307 gagcactcaccggtgaaccgctlagcccccacagttgaccgcaccacccgcgcgcgcgcgcgcgcgcgcgc 366
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17928 GAGCACTACCGGTGAACCGCTAGCCCCACACATTGACCGCCACCCGCCCGCCCAACGCC 17869
QY 367 agggcctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 426
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17868 AGGGCTGATGTCGGCGAGGCGCAGCTCAAAAGTGATTTCGCGCGCGCTTTTTCGCCACCTGCC 17809
QY 427 gccgcggtgagtggtccaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 486
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17808 GCCCGGTGGATGTCTCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 17749
QY 487 caatcgtcccgacgagctggccgctacgcccagcggtcatgagctggctacacccc 546
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17748 CAATATCGTCCGACGAGCTGGCCCGCTACGCCGAGCGGTCATGAGTGGCTACACCCC 17689
QY 547 aacgcgacctcaccgacacgaacgacgcccgcgaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 606
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17688 GACCGCGACCTACCGGACCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 17629

```

QY 607 caatcagcgcatgtcacgctgaagtgcctacctgaccccccaagcgcgggcccaccttt 666
|||||
Db 17628 CAATACAGCGCATGTACGGCTAAGTGGCTACCTGACCCCCCAAGCGCGGGCCACCTTT 17569
QY 667 gaagccgtgtctagcacaactggccgccccccggcgcgacccaacccccgcagcaccaccccg 726
|||||
Db 17568 GAAGCCGTGTAGCAAACTGGCGCGCCCGCGCGCGACCAACCCGACGACACCCCG 17509
QY 727 atcatcagacaccccccgatggcgccgcacatgacgcgcagacccccgacccgaagcccaa 786
|||||
Db 17508 GTCATCGACACACCCCGATGCGCGCGCCATCGACCGCGACACCCCGACGCAAGCCCAA 17449
QY 787 cgaacacacgagggctgctggcgggctgcgcgcgctgctgcctcccggaacatgggc 846
|||||
Db 17448 CGCAACACGACGGGTGCTGGCGGGCTGGCGCGCTGATCGCCTCCGGGAACATGGCG 17389
QY 847 caacacaaagggtcttcgctctcgatggtgttcacacacccacctgacggacctgcaaac 906
|||||
Db 17388 CAACACAAAGGTCTTCCCGTCTCGATGTGTTCACCAACACCCCTGACCGACCTGCAAAAC 17329
QY 907 ggcgcgcgaagggttcacccgcgcgccgacccctgtaccatgcccgtatgacccac 966
|||||
Db 17328 GCGCGCGCAAGGGTTCACCGCGGGCGGACCCCTGCTACCATGGCCGATGTGATCCCG 17269
QY 967 atgaccgacacgcccaccactactcccccgcgaagggaggtacccccaggcgatcttc 1026
|||||
Db 17268 ATGACGACGCGCCACCACTACTCCCGCGCAAGCGGAGGTACCCCGAGCGATCTTC 17209
QY 1027 gaccagcgcacccctggcgctgtatcacacaaagcctgaacctccccggccagcg 1086
|||||
Db 17208 GACCAGCGCACCCCTGGCGCTGTATCACCAAAAGCGCTAGCCCTCCCGGCCGACCGG 17149
QY 1087 atcatgtgttcgcaacgacggcggtgcacccaacccggctgtgacgcacccgacctac 1146
|||||
Db 17148 ATCATGTGTTCGCAACGACCGCGCTGCACCAAAACCCGGCTGTGACGACCGGCTTAC 17089
QY 1147 cagaagcaagccacacagtcacccgctgaccagcaccggaacgacacacacccgag 1206
|||||
Db 17088 CACAGCAAGCCCGACCGCTGACCGCTGACCGACCGGACCGACCGACGATCACCCGAG 17029
QY 1207 ctgacctggctggcgccccacacacgactgcgcgaagggctggaccacccacacac 1266
|||||
Db 17028 CTGACCTGGCTGGCGCCCCGACACCGGCTGCGCGAAGGCTGGACACCCACACAC 16969
QY 1267 aacacccagccacacccgaatggttaccacacccacccacctgacacatggccaccccg 1326
|||||
Db 16968 AACACCCAGCGGCACACCGAATGGTTACCAACACCCACCCGACCTCGACCGCCCAACCGTGG 16909
QY 1327 accaacaccttcacaccccc 1347
||| ||| ||| |||
Db 16908 ACCTGTGAGATACACTACAC 16888

RESULT 8
MTCY251/c
LOCUS MTCY251 38380 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 5/162.
ACCESSION Z74410 AL123456
VERSION 274410.1 GI:3261600
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 38380)
Cole, S.F., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,

Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
2 (bases 1 to 38380)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1405752.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1..38380
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
..35889
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83342"
..956
/clone="Y251"
/gene="hycB"
..956
/gene="hycB"
/note="KV0084, (MTCY251.02), len: 316, hycD, similar to
hycD_ECOLI Phd40 format; hydrogenlyase subunit 4 (307 aa)
FASTA scores, opt: 570, z score: 591.6, E(): 2.1e-26,
(14.8% identity in 405 aa overlap) and NUCL_ECOLI
hycD nadh dehydrogenase i chain h (325 aa) FASTA scores,
opt: 207, z score: 220.7, E(): 9.5e-06, (26.5% identity in
260 aa overlap)"
/codon_start=1
/transl_table=1
/product="hycB"
/protein_id="CAA98920.1"
/db_xref="GI:1405754"
/db_xref="SPTREMBL:O10881"
/translation="MSYLACAAOIGVMVCAPIVIGTQVRAWELCRAGAGLLQPWR
LLKQLGKQITPAGTIVFAAPVIVAGTITLAAIAPLAVTSGPDSADILFAVVG
LLEFLGYALTLAGIDTGTSGFGMGASRETITLAVFALISFAGSANGAL
VASTIDHPGHVSLAGVLAPVALVIVVAVTGRPDVNPATHLTMTVHEAMVLEYAG
PRALVEMAGMRLTVLLALLFLPWGIAAGAPTALDVLTCGWAAAKVA1LAVLL
ATFEVFLAKLRFVRPELLAGSFLALLAVTAANFTTGA"
967..1629
/gene="hycP"
967..1629
/gene="hycP"
/note="RV0085, (MTCY251.03), len: 220. hycP, similar to
splp77524|HYFE_ECOLI HYDROGENASE-4 COMPONENT E (216
aa), FASTA scores: opt: 204 z-score: 272.1 E(): 1.2e-07;
25.5% identity in 216 aa overlap"


```
QY 487 caatatgctccgacgaagctggcccgctacgcccagcgggtgatgactggtacacccc 546
|||||
Db 12293 CAATATGCTCCGACGAGCTGGCCCGCTAGCCCGACGGGTATGACTGGCTACACCC 12234
|||||
QY 547 gacggcagctccacgacacccgaacgcgcccgcgaacgcgagtcacccctgagcaacag 506
|||||
Db 12233 GACGGCAGCTCACGACACCCGACCGCGCCGCGCAACGGGGATCACCTGAGCAACAG 12174
|||||
QY 607 caatacagcggatgtacagggtaagtgtgctactgaccccccacgaacgcggccacttt 666
|||||
Db 12173 CAATACGACGGCATGTACCGCTTAAGTGGCTACTGTAC-CCCCAAGCGGGCCACCTTT 12114
|||||
QY 667 gaacgctgtacccaactggcccccgcggcgagcaacccccgaacacaccccg 726
|||||
Db 12113 GAACCGGTGTACCAAACTGGCCGCGCCCGCGCGGACCAACCCGACACACCCCG 12054
|||||
QY 727 gtatcgacaccccccgatgcggcgccatcgacgcgcgacacccgcagcgaagccaa 786
|||||
Db 12053 GTATCGACACACCCCGATGCGCGCGCATCGACCGCGACACCCGACGCAACGCCAA 11994
|||||
QY 787 cgaacacacgaacgggtgtgcgggggtgcgcgctgacgctccgggaactgggc 846
|||||
Db 11993 CGCAACACACGAGGGTGTGCGCGGGCTGCGCGGCTGATCGCTCGCGGAACCTGGC 11934
|||||
QY 847 caacacacggcttcccgctctcgatcggtggtcacacacacccctgacgaactcaaac 906
|||||
Db 11933 CAACAAACGGTCTCCCGCTCTCGATGCTGTGTCACCAACACCCCTGACCGACTGCA 11874
|||||
QY 907 ggcgcggcgaagggttcacccggcgccgacccctgctacccatggccgatgtacccq 966
|||||
Db 11873 GCGCGCGGAAGGGCTTACCGCGCGCGCGACCCCTGTACCCATGGCGGATGTGATCCG 11814
|||||
QY 967 atgacgacgcgcgcacactactcccccgcaagcgaggaggtacccccggagatctc 1026
|||||
Db 11813 ATGACGACGACGCGCCACCACTACTTCCCGCGCAAGCGGAGGTACCCCGGATCTTC 11754
|||||
QY 1027 gaccacggcacacccctggcgctgtatcacacacacacgcctagcctccccggccagcg 1086
|||||
Db 11753 GACGACGACACCCCTGGCGCTGTATCACACCAACACCCCTAGCCTCCCGGCCACCGG 11694
|||||
QY 1087 atcatgcttcgcaacgacggcgtgcacaaacacggcgtgtgacgacccggctac 1146
|||||
Db 11693 ATCATGCTTTCGCCACGACCGCGGCTGCACCAACCGGCTGTGACGACCGGCTAC 11634
|||||
QY 1147 cagcgaagccacacagctcacgcctggaccgacccgacacacacacacacacacac 1206
|||||
Db 11633 CACAGCAAGCCACACCTCACCGCTTGACGACGACCGGACGACGACGACGACGACGAC 11574
|||||
QY 1207 ctgacccctggcgcgcccgcgaacacgactcgcggaagagctggacacccacac 1266
|||||
Db 11573 CTGACCCCTGGCGTGGCGCCCGCAACCGGCTCGCCGCAAAAGGCTGGACGACCCCA 11514
|||||
QY 1267 aacacccagggcacacgaatggctaccacaccccccacacacacacacacacacac 1326
|||||
Db 11513 AACACCCAGCGCACACGGAATGCTACCAACACCCACCTCGACACGCGCCACCCCGC 11454
|||||
QY 1327 accaacccttcacaccccccgaacgattcctccacacacacacacacacacacac 1386
|||||
Db 11453 ACCAACACCTTCACACCCCGGACGATTCCTCCACACCAACGACGACGACGACGAC 11394
|||||
QY 1387 gattgac 1393
|||||
Db 11393 GATTGAC 11387
|||||
```

```
RESULT 9
AE006921/c
LOCUS Mycobacterium tuberculosis CDC1551, section 7 of 280 of the
DEFINITION complete genome.
ACCESSION AE006921 AE000516
VERSION AE006921.1 GI:113879142
```

```
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 9764)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Direct Submission
JOURNAL
FEATURES
source
1..9764
location/Qualifiers
organism="Mycobacterium tuberculosis CDC1551"
strain="CDC1551"
db_xref="taxon:83331"
note="clinical strain"
122..1414
gene="MT0095"
note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; similar to SP:PI6431 GR:X17506
SP:41684 PID:882614 GB:000096; identified by sequence
similarity; putative"
1449..2123
gene="MT0096"
1449..2123
gene="MT0096"
note="Identified by Glimmer2; putative"
codon_start=1
transl_table=1
product="hypothetical protein"
protein_id="AAK4439.1"
db_xref="GI:13879143"
translation="MSVTKHAPSRVRLHUTSTVVKGKSGSLSWRRVYTGDLGLAVNK;
GREYRAVKPVLPGTLQPKDMMTVTVDAGPGRVSRVEVAAPAAELPAIVADPRRIH
FLDGGSTVKGNIKVPAKLVGSKSTKMKLPGLPYRITSRTALKNELVECSHPLGH
KRWKPFESI.SPTLIRVTTFDYHAAGAATKNGLKFEYEMTGFAKSNAGIEATIAKISDU
YAKGVA"
2280..2873
gene="MT0098"
2280..2873
gene="MT0098"
note="Similar to GP:145427; identified by sequence
similarity; putative"
codon_start=1
transl_table=1
product="methyltransferase, putative"
protein_id="AAK44320.1"
db_xref="GI:13879144"
translation="MDQPNANTHYDALLDANVPLGTQCVLDVCCGGGLLAARLARI
PYTAVDIDAPVLRRAOTRFANAPIRWLHADIMTAELPNAGFDVAVSNAALRHIEDTR
TALSRLGLVTPGGTLAVVTFTPSLRNGHLHLSWVACGNANRVKQKWEHSAPIKWP
PPQTLHELRSVRALLPGACTIRLLYGRVLVTRAPV"
3002..3772
gene="MT0099"
3002..3772
gene="MT0099"
note="Identified by Glimmer2; putative"
codon_start=1
```



```
Qy 547 gacggcaccctcacgacacaaacgagccgcccgaacgaggcgatcacccctgagcaaacag 606
Db 9077 GACGGGACCTTACCGACACCGACGCGCCCGCAAGCGGCATCACCCCTGAGCAACAG 9018
Qy 607 caatacagcgcgatgtcacggctgaagtgtctacctgaccccccaagcgcgggccaccttt 666
Db 9017 CAATACGACGCGATGTACGGCTAAGTGTCTACCTGACCCCGCAAGCGCGGCGCACCTTT 8958
Qy 667 gaagcctgtctagcaaaactggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 726
Db 8957 GAAGCGGTGTAGCAAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8898
Qy 727 gtcatcgacacacccccgatggcgccgcatcgacgcgacacccgcagcgaagcccaa 786
Db 8897 GTCATCGACACACCCCGATGGCGCGCGCATCGACCGGCACACCCCGACCGCAAGCCCAA 8838
Qy 787 cgaacacacgagggctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 846
Db 8837 CGCAACACGACGGGCTGTGGCGGGCTGCGCGCGCTGATCGCGCTCCGGCAACTGGCG 8778
Qy 847 caacacacggtcttccgctcgtcgtgtgtcaccacacccctgacccacctgcaaac 906
Db 8777 CAACACACGCTTCCCGTCTCGATCGTGTGTACACACACCCCTGACCGACCTGCAAAAC 8718
Qy 907 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 966
Db 8717 GCGCGCGCAAGGGCTTACCGCGCGCGCGCACCTGTACCCATGGCGCGATGTGATCCGC 8658
Qy 967 atgacacgacacccccactactccccgcgaagcggaaggtacccccaggcgatcttc 1026
Db 8657 ATGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8598
Qy 1027 gaccacggcacacccctggcgctgtatcacaccaaaccgcctagcctccccggcccgag 1086
Db 8597 GACCACGGCACACCCCTGGCGCTGTATCACACCAACGCGCTAGCCTCCCGCGCGCAGCG 8538
Qy 1087 atcaat..tggtcccaacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1146
Db 8537 ATCATGCTGTTCGCGAACGACCGCGGCTGCACCAAAACCCGGCTGTGACGACCGGCTAC 8478
Qy 1147 cacagcaagccacacagtcacgcctggacagcacgcgcgcgcgcgcgcgcgcgcgcgcgc 1206
Db 8477 CACAGCAAGCCACACGCTGACCGCTGGACGACGCGGAGCAGCAGCAGCAGCAGCAGCAG 8418
Qy 1207 ctgacctggcctgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1266
Db 8417 CTGACCTGGCCTGGCGCGCGCGCGCACACCGACTCGCGGAAAAAGGCTGGACCAACCAAA 8358
Qy 1267 aacaccacggcacacccgaatggtaccacacccacctcgaccacgcgcgcgcgcgcgcgcgc 1326
Db 8357 AACACCCAGCGGCACACCGAATGGCTACCAACCCACCTCGACACCGGCAACCCCGC 8298
Qy 1327 accaaccttcacacccccgaagattcctccacacaaagacgacgacgacacaccc 1386
Db 8297 ACCAACCTTCCACACCGCGAAGATTCCTCCACCAACCAAGACGACGACGACGACGACGAC 8238
Qy 1387 gattgac 1393
Db 8237 GATTGAC 8231
```

```
RESULT 10
MBDR3S1
LOCUS MBDR3S1 9281 bp DNA BCT 27-APR-1996
DEFINITION Mycobacterium bovis deletion region 3, 5' end.
ACCESSION U35017
VERSION U35017.1 GI:1049238
KEYWORDS
SEGMENT 1 of 2
SOURCE Mycobacterium bovis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
```

```
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 9281)
AUTHORS Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE Molecular analysis of genetic differences between Mycobacterium
bovis BCG and virulent M. bovis
JOURNAL J. Bacteriol. 178 (5), 1274-1282 (1996)
MEDLINE 96200095
REFERENCE 2 (bases 1 to 9281)
AUTHORS Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
Pathogenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES
location/Qualifiers
1..9281
/organism="Mycobacterium bovis"
/db_xref="taxon:1765"
misc_difference replace(1400..9281,"")
BASE COUNT 1706 a 3191 c 2929 g 1453 t 2 others
ORIGIN
```

```
Query Match 92.0%; Score 1281.8; DB 1; Length 9281;
Best Local Similarity 97.2%; Pred No. 13e-165;
Matches 1304; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 7 cgggtagcgcgaacgagattgtcgaggtctttgatgc--trgatcccgagctggaccgct 66
Db 84 CGAATAGCCGGAGGAGCTTGTGAGGTTCTTGTATGCGCTGGATCCGAGCTGGACCGCT 143
Qy 67 tggacgaggtgtcttttgaaggttgaccaccccaagaacgctcggtctctggaacgtc 126
Db 144 TGGACGAGGTGTCTTTTGTAGGTGTTCACACCCCGGACGGCTCGGGTCTCTGGAACGTC 203
Qy 127 tggaaatgcttggcgccggctacacgcgcgtgggtcgcgcgttatcaacaaactgacg 186
Db 204 TGGAAATGCTTGTGCGCGGCTACCGCGGTGCGGCGACACAGTTGATCAACCAACTCGACA 263
Qy 187 cccaagccagtcuquaaactggtggtgacgctgctgtcgtcgctgcccacacggtttac 246
Db 264 CCGAAGCCAGCAGGAGGAACCTGGGCGGCGACGCTGTGTGCGGCTGCCAACCGGTTAC 323
Qy 247 gcatcacaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
Db 324 GCATCACAAGCCGACCGCCGCTTACGATCGCGCGACGCGCGGATCTCGGACTCTGTC 383
Qy 307 aadactcaatvmtgaatcgtatgacacagttgacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 366
Db 384 GAGTACTTACGATGACCGCTAGGCGGACAGTTGATCGCGCGGCGGCGGCGGCGGCGGCGG 443
Qy 367 agggcctgacggtgagcgcaagtgattcgcccttttgcgcacactgccc 426
Db 444 AGGGCTGATCGCGGAGCGGCACATCAAGTGATTGCGGCCCTTTTTCGCCACCTTGCCC 503
Qy 427 gccgcggcggaatgltccaccccgccgagccgcgaagccgcgcgcgcgcgcgcgcgcgcgcgc 486
Db 504 GCCCGGTGGATGTGTCCACCGCGCAGCGCGCGGAAAGCGACCTGGCGGCGGCGGCGGCGG 563
Qy 487 caatatgcctccgacgagctggcccgctacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 546
Db 564 CAATATGCTCCGAGAGCTGGCCCGCTACGCCGCGGGTTCATGGACTGGCTACACCCC 623
Qy 547 gacggcaccctcacgacacccgaacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 606
Db 624 GACGGCAGCTTACCGACACCGGAACGCGCGCGCAACCGCGCATCACCTGAGCAACAG 683
Qy 607 caatacagcgcgatgtcacggctgaagtgtgctacctgaccccccaagcgcgggccaccttt 666
Db 684 CAATACGACGCGATGTACGGCTAAGTGGTACTGTACCGCGCGGCGGCGGCGGCGGCGGCGG 743
Qy 667 gaaagcctgctagcacaactggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 726
```



```
gene
complement(3286. .3654)
/gene="Rv1137c"
CDS
complement(3286. .3654)
/gene="Rv1137c"
/notes="Rv1137c, (MTCI65.04c), len: 122. unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1137c"
/protein_id="CAB09028.1"
/db_xref="GI:2117175"
/db_xref="SPTREMBL:O06537"
/translation="MLSARCHIRHISPGDKARCAHLSATLRPGIGISPTNVGNATVL
ADGTPAKPIQCAETMORARHTGSCFSANARGPALSSGNPSRAGCGVSSITTSPTPQ
AIRLLACTSDALVTYTPAR"
complement(3659. .3662)
/notes="possible RBS, AGGA, for Rv1137c"
complement(3671. .4687)
/gene="Rv1138c"
complement(3671. .4687)
/gene="Rv1138c"
/notes="Rv1138c"
/notes="Rv1138c, (MTCI65.05c), len: 338. Similar to many
M.tuberculosis hypothetical proteins e.g. TR:P72008
MTCY04C12.35 (412 aa), fasta scores; opt: 89 z-score:
292.1 E(): 4.5e-09, 24.6% identity in 358 aa overlap.
TBparse score is 0.929"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1138c"
/protein_id="CAB09026.1"
/db_xref="GI:2117176"
/db_xref="SPTREMBL:O06538"
/translation="MPSYDDLLVGGPGCLATALHARAGLSIVAPRENPIDKA
CGEGLMPGGLAELTSLGVDPVGFHGIATVGEHRRVQARFTCPGRGVRRITLHAAL
AARAKEDTWRVATIQDARHVTAGVRAKWLVAADGLHSVRRRAVGIKATAGT
PRYGVHMYRLVPSDFVEVHWSRMGEAYVTPEPDLVGVAILSRQRPPLAWFPSLA
HHLQDSRGHARGCGPLRVQVSRVRAGRVLLVGDAAGYEDALTGEGISLAVKQAAAV
SAIVDDTPASVAAWHRIITRDYRLVTRGLVLAFTPRARRAIVPLCALLTAFRYGVN
ILAY"
complement(4684. .5184)
/gene="Rv1139c"
complement(4684. .5184)
/gene="Rv1139c"
/notes="Rv1139c, (MTCI65.06c), len: 166. Highly similar
to hypothetical B. subtilis protein, YRPQ. FASTA scores:
YBPQ_BACSU HYPOHETICAL 19.3 KD PROTEIN IN BC5A-DE (168
aa) opt:446; E(): 2.2e-26; 38.4% identity in 164 aa
overlap. TBparescore is 0.911"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1139c"
/protein_id="CAB09027.1"
/db_xref="GI:2117177"
/db_xref="SPTREMBL:O06539"
/translation="MYLLILAVFERLAELVVAQRNARMSFAQGGKFGFRPHYVVMV
ILHTALLGCVPEPMLHPFPFPLWGLMPLVAVASOGLRWCMVKSLGRKRWTRIVL
PHATLVRRGPRMWRHPNVAVVAEGFALPLVHTAWLTALVFTLANATLLTVLRIVEN
SVLGYT"
complement(5192. .5197)
/notes="possible RBS, AGGTGG, for Rv1139c"
5546. .6394
/gene="Rv1140"
5546. .6394
/gene="Rv1140"
/notes="Rv1140"
/notes="Rv1140, (MTCI65.07), len: 282. Weak similarity in
C-terminus to hypothetical E. coli proteins YPRa and YPRB,
possibly membrane-bound. FASTA scores: YPRa_ECOLI
HYPOHETICAL 24.3 KD PROTEIN (URF 1). - (217 aa) opt:
166; E(): 0.00062; 31.0% identity in 158 aa overlap.
TBparse score is 0.927"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1140"
```

```
/protein_id="CAB09028.1"
/db_xref="GI:2117178"
/db_xref="SPTREMBL:O06540"

Query Match      50.2%; Score 699.6; DB 1; Length 34331;
Best Local Similarity 70.4%; Pred. No. 7e-87;
Matches 975; Conservative 0; Mismatches 384; Indels 25; Gaps 2;

Qy 1 tgggttgcgggtagccgcgaacyygatlgtgcgaggtcttltgagcgtggatgcgcgagctgg 60
Db 15862 TGAGTGTGGACACACGCGCAGCAGAGATCTCCCGCGCGTGTGGATGCGCTACACAGCCTGTTGT 15803

Qy 61 acqcttggagcayugtcttlttgaagtgltlgaacaccccaagaacgagctgcgtg 120
Db 15802 CGCGGGTGCTCGATCTCAAGTCCGATGCGGTTCACCAACCCCGGAATTGCTGGCCCTGTTTGC 15743

Qy 121 aaqcttctggaatgcttltggttcyccggttaccgcggglgggtcacgcgttgaataaaccaac 180
Db 15742 AGCGACTCGAGCTTCGAAAGGGCGCGCCAGGTCGCGGTTCGACGACGCGCTTGTATCAACCAAC 15683

Qy 181 ttgacgcctaaagcagcgaggaanaaactggcgggcacagctgtgtgctgcgcgtgcgcgaacc 240
Db 15682 TCGCTGGGCAAGCTTCGCGAGCAGAGAGCTTCGCGGGAGCGCTGCGCAGCGCTTGCCCAACC 15623

Qy 241 ggttacgcatacaaaacccgcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 15622 GGCTACACATCACATCCCGGTGAGGCCAGCGCGCATTCGCCGAAGCGGAAGACCTCGGTTG 15563

Qy 301 ctgcgtcagcacttaccggtgaacgcgttagcccaagttgacgcgaacccgcgcgcgcgcgcgcgc 360
Db 15562 AGCGCGCGCCCTCACCGGTGAACCGTTCGCGAGCGCAGTTGACCGCGACCGCGCGCGCCGCTC 15503

Qy 361 aacgcgcagggccttactcgcgcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 15502 AACGTGAGCGCAAGATCGCGCGCAGAACACATTAAGAGAGATCCAGG-CCTTCTTCAAGCAG 15444

Qy 421 ctccgcctccatcaatgagtgatgatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
Db 15443 TTGTTCGCCCGCGGTGATCTGGGTATCCGCGAGCGCCCGGAGGCCAGCTGGCGCGAACTG 15384

Qy 481 gccgcctcaatctatccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 15383 GCCACCACTCGATCTCCCGATCACTTGCATTCGCTGCGCAGCGAGTGTGATGGACTGGCTG 15324

Qy 541 cacccttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 15324 CACCTTCGATCTCAAGTTTTCGATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 15264

Qy 601 aaccagcaatcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660
Db 15263 AACGACGAATTCACCGGATGTTCACGTATCAGCGGTCTGCTGACCCCGGAGTTTGGCGGCC 15204

Qy 661 accttgaagccglgctagccaaactggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 15203 ACCATGAGCGCGTGTGTGGCCAAACTTGGCCGCACCGCGGGCGTGTGCAACCCCGATGACCAG 15144

Qy 721 accccggtcatctaaacacaccccgatctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 15143 ACCCCGCTCGTGTGATGACACACCGGATTCGCGACGCGGTGCGCGCGCACACCCCGCAGCCAG 15084

Qy 781 gcccaacgcaaacacacagcggtgctlqgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 15083 GCGCAACGCAATCATGATGCGTTTTTGGCGCGCGCTCGCGGGCCCTGTGGCTCCCGTGAG 15024

Qy 841 ctgggccaacacacgggttctccgtctcgatcgtggtcacccaccacctgacccgacctg 900
Db 15023 CTGGGCAACACAAAGGGGTGCGGTAAACGATCGTGTGTGAGCAGCAGCTAAGAGAGCTG 14964

Qy 901 caaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 14963 GAAGCCCGCACCGCAAGGGGTAAACGGTGTGTGTGCGGGGTGCGGATGTCGGACCTT 14904

Qy 961 atccgcgatgaccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
```

Db	14903	ATCCGGATGCGAGCCACGCAACCACTA-----TCTGGCA	14868
Qy	1021	atcttcgacacgacacacccctggcctgtatcacacaaacgacctgaccccgcc	1080
Db	14867	TTGTTTACCGCGGTAGCCGTTGGCTGTGTATCATCACCAAGCGGTAGTTCCTCCCGCGG	14808
Qy	1081	cagcggatcatgtcttgcgcaacgacgagcggtgtgacacaaacccgctgtgacgacccg	1140
Db	14807	CAGCGAATCATGTTGTACGGAAGGATCGTGGCTGTCTCCAGGCGGGTTCGACGCCCCG	14748
Qy	1141	gcttaccacagcaagccacccagctcacgcgctggaccagcaccgagcagcagacalc	1200
Db	14747	GCCTTACCACATGAGGTCTCCACCGTAACCGCTGGTGAACACCAACCCACCGTACCGACATC	14688
Qy	1201	accgagctgacctggcctggcgcgcgcgacacccgactgcgcgaacgaagctgacacac	1260
Db	14687	AACGACCTACGCTGGCTGGCGCCGCGACATCGCTTGTCCAAAGGCTCGAAGACC	14628
Qy	1261	cacacaaacaccccgccacacacgaatgctaccacacccacccacctcgacacgcccac	1320
Db	14627	CGCAAAACGCCACCGCGACACCGAATGGCTACCAACACACACCTCGACACCGCCAA	14568
Qy	1321	cccgcacacacactctccaccaccccgacgactctccacaaacgaacacgacgacac	1380
Db	14567	CCGGAATAATCGATACCAACCAACCCCGCAAAATCTGTGCGAACAGACGACGACGAA	14508
Qy	1381	aaac 1384	
Db	14507	CCAC 14504	
RESULT	13		
LOCUS	AE006996/c	14953 bp	DNA
DEFINITION	Mycobacterium tuberculosis CDC1551, BCT section 82 of 280 of the complete genome.		
ACCESSION	AE006996		
VERSION	AE000516		
KEYWORDS	AE006996.1	GI:13880759	
SOURCE	Mycobacterium tuberculosis CDC1551.		
ORGANISM	Mycobacterium tuberculosis CDC1551.		
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
AUTHORS	1 (bases 1 to 14953) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.		
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 14953) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1..14953 /organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" complement(172..390) /gene="MT1172.1" complement(172..390) /gene="MT1172.1"		
gene			
CDS			

/note="identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAK45431.1"	
/db_xref="GI:13880760"	
/translation="MPAGPGFTGPGVGPARGGVTSRHPALPVLTVQVYEPAPLVALLT	
TIRKASNLARRRYGPWNGPAMAVPIARI"	
314..1167	
/gene="MT1173"	
315..1167	
/gene="MT1173"	
/note="identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAK45432.1"	
/db_xref="GI:13880761"	
/translation="MPDYTPARWAHAWAGPRPARMHPANQAPARHDSHNKSPACMS	
QSTPYRSSLAEFRRAITNVAVPHIEPPGIVRRRRVVGVTVLIGAVML;FSLRRTP	
GKSSFYMLTIALAAVWIAAGALMSGPLHGGICWRGNORPVI.TGTTVCLLAGIFGVG	
AMIVRAIPCAEPIAKVLOFAHOCTLLPILITLLINGIAEEMFRGALYTAIGRRYPV	
TLSPIVYVCATMASANI.MIGFAAIFVGTVCALERRASGGVLAFLTHFVWGLIMVFAI.	
PLFAV"	
complement(1175..1981)	
/gene="MT1174"	
complement(1175..1981)	
/gene="MT1174"	
/note="similar to GB:U00096 PID:1742270 PID:1742276	
PID:1787660; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="enoyl-CoA hydratase/isomerase family protein"	
/protein_id="AAK45433.1"	
/db_xref="GI:13880762"	
/translation="MPPSGIAALTPVTLNVTLDRLVLSVRINRPSLSNSI.TVPILTG	
IAUTLERAADPPVKVYKRGVGRGSGVSNVDVWGGPPTAIVEENRAVRAAA	
ALPHPVAVVVGPGAVGVASLACDFILASDSAPFMLANTKVAMPDGGASALVAAA	
TGTRAMRLALAEQIPAREALWGLISAVYPSDSDFEAEVDKVISRI.LAGPAIAFAQA	
KNAINAAALTELRPTFAFPLDQGEVILLRTHDFAECAAAFLQRRTPNF.TGS"	
complement(2124..2930)	
/gene="MT1175"	
complement(2124..2930)	
/gene="MT1175"	
/note="similar to GB:U00096 PID:1742270 PID:1742276	
PID:1787660; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="enoyl-CoA hydratase/isomerase family protein"	
/protein_id="AAK45434.1"	
/db_xref="GI:13880763"	
/translation="MSNYKRIIDTRTVPLGLAVTLADGVLSVTIDRPESINSI.TKPIVLG	
MAAIEGAATDPVKVYKRGVGRGSGGAI.SYDDVWASGPTDTVAEANRVRAIV	
AI.PQPVAVVVGPTVCGVSLACDLVLSADNAFFMLHTNVGLMPDGGASALVAAA	
IGTRAMHMAILPKRPPAALSMGLSVSAVYPAADFAEVDKLI.SRI.LAGPAIAIAKT	
KNATNAATLTLEAPTLLELDGQALLRTDDFABGATAFOQRRTPMF.TGR"	
2929..4116	
/gene="MT1176"	
2929..4116	
/gene="MT1176"	
/note="similar to GP:4539189; identified by sequence	
similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="fatty acid-CoA racemase"	
/protein_id="AAK45435.1"	
/db_xref="GI:13880764"	
/translation="MWTLTIDDAVTELRHDLRRYAVEGVNGRRKAGTFMAGPLSLGR	
VPELAGIGPGPHAAAILLGDLCADVIRDRPSVDGISRDAMLNRRIVTAU.KSDQGI.	
EIALKLIKADVLIEGYRPGVTERIGLGPCEAKVNDRLIYARWTGWCOTG;PRSQOGL	
HDINYLSLNGILHAIGRDERPPVPLNLVGDGCGSMFLVGLI.LAAL.WERJSSKQGV	
VIAAMVDGSSVLIQMMWRATGNWTDTRCANMLDGGAPPYDTIECAIGRYVAVGATV	
IQFYAAMLACGLDAELPQNDRAKPEURLALLTEAFASIDRHDHMCVAVFANSICVCT	

Db 9856 GCACACGCAATCATGATCGTTTTTGGCCGCGTGCAGGCGCTGTTCGCCCTCCGGTCAG 9797
 QY 841 ctggggccacacacaggtcttcctccgtctcgatcgtggtgcacacacacccctgacgcactg 900
 Db 9796 CTGGGCCAACACAGAGGGCTGCGGTAAACGATCGTGTGTGACACACACACCTAAAGGAGCTG 9737
 QY 901 caaacggcgccggcgaagggtctccacggcgcgccgacccctgctaccacatgcccgtgtg 960
 Db 9736 GAAGCGCCACCGCAAGGGGTAAACGGTGTGTTCCGGGTGTCCTGACGACCTT 9677
 QY 961 atccgcatgacacgacacgcccacacactactcccccgcaagcgaggtacccccagcg 1020
 Db 9676 ATCCG:ATGCGGAGCACGCCCAACCACTA-----TCTGCCA 9641
 QY 1021 attctcgacacggcacacccctgctgctgtatcacaccaaaccctagctctccccggcc 1080
 Db 9640 TTGTTTACGGCGCTAAGCCCTTGGCGTGTGTATCACCAAGCGGTAGCTTCCCGCGCG 9581
 QY 1081 cagcggtatcgtgttgcgcaacgacggcggtgcaccaaaccggtcgtgtagcgaccg 1140
 Db 9580 CAGCGAATCATGTTGATACGCGAAGGATCGTGGCTGCTCCAGGCGGCTTCCGACGCCCG 9521
 QY 1141 gcttaccacacgaagccacacgctcacgcgctgacgcgctggccgacccgacgcacgacatc 1200
 Db 9520 GCCTACCACTAGTGGTCCACCGCTAACCGCTGGGACACCAACCAACCGCTACCGACATC 9461
 QY 1201 accgagctgacccctggcgtcgccgcgacacgcgactcgccgaagagctgacacacc 1260
 Db 9460 AACGACCTACGCTGGCTCGCGCCCGACCAATCGCTTGTCCGAAAGAGCTCGAAGACC 9401
 QY 1261 cacacacacacacgacacacgaatggtctaccacacacccacccctcgaccacggccaa 1320
 Db 9400 CGCAAAACGCCCGACCGGACACCGAATGGCTACCAACCAACCGCTCGACACCGCCAA 9341
 QY 1321 cccgcgaccacacactctccaccaccccggaacgacttctccacacaaacgacgacgac 1380
 Db 9340 CCGCAATAAATCGATACCAACCGCCCGCAAAATCCTGTGCGCAACAGACGACGAA 9281
 QY 1381 aaac 1384
 Db 9280 CCAC 9277

 RESULT 14
 MTCY9F9
 LOCUS MTCY9F9 33100 bp DNA BCT 03-AUG-2001
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 87/162.
 ACCESSION Z84498 AL123456
 VERSION Z84498.1 GI:3261701
 KEYWORDS
 SOURCE Mycobacterium tuberculosis H37Rv.
 ORGANISM Mycobacterium tuberculosis H37Rv
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 REFERENCE
 AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
 Tekla, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrell, B.G.
 TITLE Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 JOURNAL Nature 393 (6685), 537-544 (1998)
 MEDLINE 98295987
 REFERENCE 2 (bases 1 to 33100)
 AUTHORS Parkhill, J.
 TITLE Direct Submission

JOURNAL

COMMENT

FEATURES

source

source

gene

CDS

misc_feature

gene

CDS

Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1806198.

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

Location/Qualifiers

1..33100
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 <1..32559
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 /clone="Y9F9"
 265..1380
 /gene="Rv1922"
 265..1380
 /gene="Rv1922"
 /note="Rv1922, (MTCY09F9_42c), len: 371 aa. Similar to penicillin binding proteins and eg DAC_SIRSQ P15555 d-alanyl-d-alanine carboxypeptidase (406 aa), fasta scores, opt: 382.E(); 3.1e-17, (28.0% identity in 379 aa overlap); containsPS00013 Prokaryotic membrane lipoprotein lipid attachment site"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1922"
 /protein_id="CA06504.1"
 /db_xref="GI:1806240"
 /db_xref="SPTREML:P95291"
 /translation "MISTVTFASIRRMRLGLIAATLLGCGCTGQHTTAASTVTPHIK
 ASSQVILKAIINAEKCTSAAGVKKVWVGVRGIADLASGAKLTDTDFDIASVSK
 QFTATALLVAEAKLTIDPISQVYPELPDMQATVTVEQLMIQTSGITPDYVALLAAR
 GYVSDRTIKAEAKALAAPELOFKPGRFDYNSNYLLGELIHRASGQPLPEFLS
 AEIPLGLANVYDPGVKPNKAVSYEKGTGGRNRYGAGIYSRADTLDHAGAWAGF
 IARWADNRTGSGVGIKLLAQLAGAVETEPGGDRYGAGIYSRADTLDHAGAWAGF
 VAFHISDRTSTVAISCNIDKPDVPVAMADALGRLLM"
 307..339
 /gene="Rv1922"
 /note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
 1371..2711
 /gene="lipD"
 1371..2711
 /gene="lipD"
 /note="Rv1923, (MTCY09F9_41c), len: 446 aa. Similar to esterases, beta-lactamases, eg Gl51214 esterase, (389 aa), fasta scores, opt: 569, E(); 5.4e-29, (33.7% identity in 401 aa overlap) also MTCY277.19 (56.6% identity in 410 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="lipD"
 /protein_id="CA06503.1"
 /db_xref="GI:1806239"

```

/db_xref="SPTREMBL:P95290"
/translation="MDVAGLPRLAGTQAIIHGMAQPSLLITDNLGPFVQVQACDS
RFTGVIIRAFAGLPFGFGGALSVIDGRQYVDVMTGSDRQGVPMPTADTGAMVFS
ATGLAATVIRHLVDRLGLSDYDPAVEAYWPEFGANGKSEVTVSDVLRHSGLAHLKGV
DKDEVMDHLMQKLAAPLDROHGKLAYHAVTYGWLGLSLARAYTGKGMRELFRLEL
ARPLNTDGHLAGRPADSPTKAAQTLLPOAKVPTPLDFIAPKVAGLSFGSLGLGAVYF
PGTLLSQDDMPFLDGEVPAVNGVVTARALAKTYGALANDGVIDGTRLLSSQAVRGLT
GKSELMPDLNLGLPTYGHOGYSSVPVGLLEGYGHIGLGTIGWADPETGSAFGVHN
RLTLILLDIFGSFAGLAALLNSAVVAARDPDLPLEVPHFGAPYSEPRHBOAASGA"
complement(2748..3128)
/gene="Rv1924c"
CDS
complement(2748..3128)
/gene="Rv1924c"
/notes="Rv1924c, (MTCY09F9.40), len: 126 aa. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1924c"
/protein_id="CAB06502.1"
/db_xref="GI:1806238"
/db_xref="SPTREMBL:P95289"
/translation="NDPADVNLNPTSTRDAALRVLAYRQVRARPLLIRATLAVVGGG
LFVSLPMTVLPLPELGIPALLVAFRLLAVEAQAARAYATWDRFTQUREFWHQVLV
TRAAILVGLFLAAVALVWLLVTEF"
3285..5147
/gene="fadD31"
CDS
3285..5147
/gene="fadD31"
/notes="Rv1925, (MTCY09F9.39c), len: 620 aa. fadD31
Acyl-CoA synthase similar to eg Y06J_MYCTU Q10976
hypothetical 67.9 kd protein cy338.19 (626 aa), fasta
scores, opt: 945, E(): 0, (39.8% identity in 598 aa
overlap), also G1171128 SAFRAMYCIN MX1 SYNTHETASE B(1770
aa), opt: 845, E(): 0, (37.4% identity in 593 aa overlap)
also MTCY19C5.07 (36.9% identity in 590 aa overlap). Also
similar to Rv3801c, (637 aa, 47.1% identity in 612 aa
overlap)"
/codon_start=1
/transl_table=11
/product="fadD31"
/protein_id="CAB06501.1"
/db_xref="GI:1806237"
/db_xref="SPTREMBL:P95288"
/translation="MNGSRQELRVRSGLLQIEDCLDADGGIALPAGTTILSLITERNI
KYDGVYAIYLDHARSAAGCALEVTWQGMRLAATGAHVORFAGPGDRVAILAPOG
IDYVGGFYAATKATVAVPLFAPELPGHAEIRDLTALRDSAPAVILLTAAAKNAVEGFL
NNVPLRKPTLVLDIQIPDREGELFVPEMDIDAVSHLOVTSRTPPVGVETHRVPL
GTNLVQMLISDLNRRNTHGVSWLPLVHDMLSMIGHFPVAVYGGHSTLMSPTAFVRPL
RWIQLSEGSTGRVTTAAPNFAVEAAORCLPAGQUDVDLSNVVLIIGSEPVSIDAV
TTENKAPYGLPRTAFKPSYTAETLLVATIDHAAEPTVYVLDIPQLGAGHATRAVA
PDANAVVHSCGHVARSNAWIVDPDTGPEAGAEPLDGEIENVWLOGDNVARYWGR
PEETRMFTGARLOSPLAEGSHADGSAIDDTWLRGTIDIGYVLDGELYITGRIDALLTID
GRNHYPQDIEATAEAASPMVRRGYITAFVTPASDGDPRNORLVI IABRAAGTSRSDPR
PALDAIRAAVCNRRHGLSVADLSFLPAGAIPTTTSGKLARQACRAQVLSGLRGVH"
complement(5155..5634)
/gene="Rv1926c"
CDS
complement(5155..5634)
/gene="Rv1926c"
/notes="Rv1926c, MTCY09F9.38, len: 159 aa. Unknown,
probably exported protein contains N-terminal signal
sequence."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1926c"
/protein_id="CAB06500.1"
/db_xref="GI:1806236"
/db_xref="SWISS-PROT:P97175"
/translation="MKLTMTKTAVALVAMAATFAAPVALAAYPIITGKLSGELTWT
DTVGQVVLGHKVSDDLKSSSTAVIPGYPVAGQVWEATVAINRIGSVTPAVSQFNARTAD
GINRVLWQNAAGPDTISGATIPQGEOSTGIYFDVTPGSPVITVAMNNGMEDLLIWEF#
5871..6644
/gene="Rv1927"
CDS
5871..6644
/gene="Rv1927"

```

```

/notes="Rv1927, (MTCY09F9.37c), len: 257 aa, Slight
similarity with Y01F_BACSU P54543 hypothetical 23.9 kd
protein in glng- (209 aa), fasta scores, opt: 230, E():
2.8e-08, (28.0% identity in 164 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1927"
/protein_id="CAB06499.1"
/db_xref="GI:1806235"
/db_xref="SPTREMBL:P95287"
/translation="MTAIPGSGAEPGESRALAGYPVTPALPRPVIFDQRWDTLTFI
IMPVLPESVAGSYPPGTRPDVADMTVGLVFPFMSSTKLGTALPIYVTPETPNV
KLYSIDNAGRHGVLFRLSELTARLVVPLTRIGLIPYAMRMRSRGGKHITVHSVRR
WPRGLRSLTITITGDLVEPTPLEVLTARWAGHTRKAGRTWVVPNEIKPMLRAAEI
AKLNDELIDASGVQPTGDIRALFSPGVHARFGRPCVVQ"
complement(6648..7415)
/gene="Rv1928c"
CDS
complement(6648..7415)
/gene="Rv1928c"
/notes="Rv1928c, (MTCY09F9.36), len: 255 aa. Similar to eg
HIDHA_ECOLI P25529 7-alpha-hydroxysteroid dehydrogenase
(255 aa), fasta scores, opt: 541, E(): 1.2e-27, (37.5%
identity in 251 aa overlap); also similar to many
mycobacterial proteins, e.g M. tuberculosis Rv1350,
Rv0927c, Rv2002, Rv2769, Rv2766c, etc; contains PS00061
Short-chain alcohol dehydrogenase family signature"
/codon_start=1

```

```

Query Match      49.1%   Score 683.6;   DB 1;   Length 33100;
Best Local Similarity 69.7%;   Pred. No. 1.1e-84;
Matches 965;   Conservative 0;   Mismatches 394;   Indels 25;   Gaps 2;

Qy 1  tgggttcgggttagcgcgcgaacgattctcgaggctctlttgatcgctggtgagctgagctgg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22188 TGAGTGTGGACACCGCGGAGAGATCTCCGCGGCTTGGATGCTTACACGCGCTCGTGTG 22247

Qy 61  acccttggcgaugtgcttlttgagctatlgacacccccacagacgctgctgctctg 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22428 CGCGGGTCTCGATCTCAAGTGCAGTTCGCTTGACACCCCGGAAATTCGTGCGCTGTGTC 22307

Qy 121  aacgtctggaatgcttggtgcgcgctacgcgcggtggctcagcgtgtgatacaacac 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22308 AGCGACTCGAGGTGCGAACGCGCGCGCCAGGCGCGCCGCGAGCAGCTTGTATCAACCAAC 22367

Qy 181  ttgacgccccaaacagcagcagaagaactggcgcgacgctgtgctgctgctgctggccaacc 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22368 TCGCTGGCCCAAAATCTCGAGTAAGAGCTCGCTCGCGGAAGCTCGCGCAGCGCTTGCCTCAACC 22427

Qy 241  qqltarcacatcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22428 GGCTACACATCATCTCCGGTGAAGCCAGCCCGCGATCGCGCGAAGCGGAGACCTCGGTG 22487

Qy 301  ctgcgtcagcactcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22488 AGCGCGCGCGCTTGACCGGTGAACCGCTGCCAGCGAGTTGACCGGACCGCGCGCGCTC 22547

Qy 361  aacgcccaggcctgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22548 AACGTGAGGCAAGATTCGCGCGGAGAACACATTAAGAGAGATCCAGG-CCTTCTTCAAGGAG 22606

Qy 421  ctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22607 TTGTCCCGCGCGTGTGATCTGGGTATCTCCGGAGGCGCGCGAGCGCCAGCTGCGCCAACTG 22666

Qy 481  gcgcgtcaatatgctccccacgagctggccgcctacgcgcagcgggtcagctgactggccta 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22667 GCCACAGTCGCGCTCCGATACCTGCTGATGGCTGCGCCACGACGCTGACTGGCTG 22726

Qy 541  caccgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22727 CACCCCGACGGCAACTTTTCCGACACGAGGAGCGTGCCCGCAAGCGCGCATCATGATGGT 22786

Qy 601  aaccagcaatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660

```



```
Db 13128 TTGTCGCGCGGTGGATCTGGGTATCCGCGAGGCGCGGAGGCCAGCTGGCGGAAC TG 13187
QY 481 gccgcctaataatgtcccgacgagactggccgcgtacgcccagcggtcatggaactggcta 540
Db 13188 GCCACCACTCGCGCTCCCGATCACGTGCTGCGCTGCGCCACCGACGCTGATGAGCTGGCTG 13247
QY 541 cccccgagcgacccctcaccgacacgaaacgacgcccgcgcaaacgcggcatcacccctgagc 600
Db 13248 CACCCCGACGCAACTTTTCGACACGAGGAGGTGCCCGCAAGCGGCGATCACGATGGGT 13307
QY 601 aaccagcaatacagcggtatgtcagcgctaaagtgtggtacctgaccccccaagcgcgggcc 660
Db 13308 AAGCAGGAATTTGACGGGATCTACGTATCAGCGTCTGCTGACCCCGGAGTTGCGGCC 13367
QY 661 accttgaagcgtgtcagcaactggccgcgtcccccggcgacgaacaaaccccgacgaccac 720
Db 13368 ACCATCGAGGGGGGTGTGGCCAAACTGGCCGACCGGGGGGTGCAACCCCGATGACCCAG 13427
QY 721 accccgggtcatcgacacacccccgatgcgccgccaatcgaccgcgacaccccgagccaa 780
Db 13428 ACCCGGTCGTGGATGACACACACCGGATGCGGACGCGGTGCCGCGACACCCCGCAGCCAA 13487
QY 781 gcccaacgcaaacacgagcggtcgtgcccgggtgctgctgctgctgctgctgctgctgctgctg 840
Db 13488 GCGCAACGACACCATGACGGTTTACTGGCGGGGTGCGCGGGTGTGTTGGCTCCCGTCAG 13547
QY 841 ctgggccaacaacagcgtctcccgctcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 900
Db 13548 CTAGGGGAGCATCGGGGTTCCCGGTGACCGTCTGTGTGTGACACACCGCTTAAAGAGCTG 13607
QY 901 caaacggcgccgcaagggttcaccgycggcgccaccctgtaccctatggccgcatgtg 960
Db 13608 GAAGCGCCACCGGCAAGGGGTAAACCGTGTGTTGCGGGTGCGGATGTCGGACCTT 13667
QY 961 atccgcatgaccagcaccacccactactccccgcgaagcgggaggtaccccccaaqcg 1020
Db 13668 ATCGGATGGCGAGCAAGCGCACCACTA-----TCTGGCA 13703
QY 1021 atcttcgaccagcgacacccctgctgctgtatcacacaaacgcttagcctccccqccc 1080
Db 13704 TGTGTTGACGGCGTTAAGCCGTTGGCATTGTATCACACCAAGCGTTAGCTTCCCGCGCG 13763
QY 1081 cagcggtatcgtgttgcgaacgacgcggtcgtgcacccaaccccgctgtgacgcaccg 1140
Db 13764 CAGCGAATCATGTTGTACGCCAAGGATCGTGGCTGCTCCAGGCGGGTTGCGACGCCCG 13823
QY 1141 gcctaccacagccaagccccaccagtcaccgccttgagaccagcagcggtcgtgatatc 1200
Db 13824 GCCTACCAACAGTGGGTCCACACGTAACGCCGTGGACAAACCCCAACGTAACGATC 13883
QY 1201 accgagctgacctggcctgcgcccgcgaacacgactcgcgcaaaaggctggaccacc 1260
Db 13884 AACGACCTCAGCTGGCTGGCGCCCGGACAAATCGCCTTGCGAAAAGGCTGGAAACC 13943
QY 1261 cacaacaaccccgcgccacacgaatggtctacacacccccaccctcgcacacgacgcaa 1320
Db 13944 CGCAAGAAGCGCAAGGGCGACACTGAATGGCTACCGCGGGCCCACTTGGACCATGGCCAA 14003
QY 1321 cccgcacccaacaccttcacaccccggaacgattcctccaacacgaagacgacgac 1380
Db 14004 CCACGATCAATCGATACCAACCCCGGAGAAATCTCTGCGAACCCGCGAGCAGCA 14063
QY 1381 aaac 1384
Db 14064 CCAC 14067
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 20:15:40 ; Search time 366.19 Seconds
(without alignments)
3261.296 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcgaa.....cgacgacaaacccgattgac 1393

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1202.4	86.3	12412	17 AAT33537	BCG deletion regio
2	455	32.7	1535	16 AAQ89200	Mycobacterium tube
3	455	32.7	1535	17 AAT33656	M. tuberculosis ce
4	455	32.7	1535	19 AAV18647	DNA for M. tubercu
5	396	28.4	650	17 AAT33658	M. tuberculosis ma
6	396	28.4	650	19 AAV18649	DNA for M. tubercu
7	94.4	6.8	30690	21 AA52301	S. avermitilis ave
8	89.8	6.4	1337	20 AA217263	Human gene express
9	89.4	6.4	12381	21 AA258381	Streptomyces averm
10	87.4	6.3	4466	21 AAA14663	Nucleotide sequenc
11	87.4	6.3	4478	21 AAA14661	Nucleotide sequenc

12	87.4	6.3	4547	21 AAA14664	Nucleotide sequenc
13	87.4	6.3	4571	21 AAA14662	Nucleotide sequenc
14	87.4	6.3	77536	21 AAA14651	Nucleotide sequenc
15	86.6	6.2	31422	21 AAA92302	S. avermitilis ave
c 16	82.6	5.9	77536	21 AAA14651	Nucleotide sequenc
c 17	82.6	5.9	114955	20 AAX53491	Human adenosine A1
c 18	82	5.9	58857	21 AAA58471	Nucleotide sequenc
c 19	81	5.8	1000	21 AAA02484	Human colon cancer
c 20	77.2	5.5	1218	21 AAA02488	Human colon cancer
21	75.6	5.4	4849	22 AAF25795	S. chrysomallus ac
22	72	5.2	1359	20 AA217254	Human gene express
23	71.6	5.1	1635	22 AA122780	Probe #12713 for g
24	71.6	5.1	1635	22 AA148082	Probe #16768 used
25	71.6	5.1	1635	22 AA108454	Probe #8445 used t
26	71.6	5.1	1973	22 AA113556	Probe #3489 for ge
27	71.6	5.1	1973	22 AA134918	Probe #3604 used t
28	71.6	5.1	1973	22 AA101446	Probe #3437 used t
29	70.4	5.1	1401	22 AAH68136	C glutamicum codin
30	70.4	5.1	309400	22 AAH68534	C glutamicum codin
31	70.2	5.0	15720	22 AA129504	C899p determined c
32	70	5.0	925	20 AAV99229	DNA encoding an ac
33	70	5.0	925	21 AAA72501	LigAT2 domain from
c 34	69.2	5.0	1127	21 AAA02477	Human colon cancer
c 35	69	5.0	1459	21 AAA02528	Human colon cancer
36	69	5.0	4674	21 AAA14666	Nucleotide sequenc
37	69	5.0	4725	21 AAA14665	Nucleotide sequenc
38	69	5.0	4737	21 AAA14668	Nucleotide sequenc
39	69	5.0	4767	21 AAA14667	Nucleotide sequenc
40	69	5.0	4818	21 AAA14669	Nucleotide sequenc
c 41	68.8	4.9	1593	21 AAA02504	Human colon cancer
c 42	68.2	4.9	1075	22 AA121781	Probe #11714 for g
c 43	68.2	4.9	1075	22 AA147064	Probe #15750 used
c 44	68.2	4.9	1075	22 AA107466	Probe #7457 used t
c 45	68.2	4.9	1403	22 AA112593	Probe #2526 for ge

ALIGNMENTS

RESULT 1
AAT33537
ID AAT33537 standard; DNA; 12412 BP.
XX
AC AAT33537;
XX
DT 15-FEB-1998 (first entry)
XX
DE BCG deletion region 3 and flanking sequences.
XX
KW BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
XX
OS Mycobacterium bovis strain BCG.
XX
FH Key Location/Qualifiers
FT misc_feature 1406..10673
FT /*tag=a
FT /*note="BCG delta 1 deletion region"
XX
PN WO9625519-A1.
XX
PD 22-AUG-1996.
XX
PF 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahairas GG, Slover CK;
XX
DR WPI; 1996-393419/39.
XX


```
XX PR 02-SEP-1993; 93US-0118442.
XX PA (CORR ) CORNELL RES FOUND INC.
XX PI Riley LW;
XX DR WPI; 1995-115442/15.
XX DR P-PSDB; AAR71931.
XX XX
XX PT DNA encoding for cellular uptake of Mycobacterium tuberculosis -
XX PT used to develop prods for vaccines, passive immunisation and
XX PT diagnosis and cellular uptake of other materials
XX PS Claim 2; Page 9-11; 46pp; English.
XX XX
XX CC The isolated DNA molecule of the invention confers on M.
XX CC tuberculosis an ability to enter cells and to survive within
XX CC macrophages. It encodes a polypeptide having a mol. wt. of about 50-
XX CC 55 kDa, pref. 52 kDa. The AA sequence represents a highly
XX CC hydrophilic protein with a hydrophobic region at its carboxy
XX CC terminus. It could be a secreted protein, a cytoplasmic protein, or
XX CC a surface protein with its carboxy terminus attached to the outer
XX CC membrane of the organism. The deduced AA sequence is in AAR71931.
XX XX
XX SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 32.7%; Score 455; DB 16; Length 1535;
Best Local Similarity 100.0%; Pred. No. 4.1e-75;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 gatcgagagcgacgtaaaagtattcgcccttttgcacactgcccgcgcg 433
DB 827 gatcgagagcgacgtcaaaagtattcgcccttttgcacactgcccgcgcg 886
QY 434 tggatgtgtccaccgagcgccgaaagcgacgtggtcagctggtacacccgacgcg 493
DB 887 tggatgtgtccaccgagcgccgaaagcgacgtggtcagctggtacacccgacgcg 946
QY 494 gtcccagacgagctggcccgctacgcccagcggtcatggactggctacacccgacgcg 553
DB 947 gtcccagacgagctggcccgctacgcccagcggtcatggactggctacacccgacgcg 1006
QY 554 acctaccacacacgacgacgcccgaacgacgtggtcagctggtacacccgacgcg 613
DB 1007 acctaccacacgacgacgcccgaacgacgtggtcagctggtacacccgacgcg 1066
QY 614 acggcatgtcagcgtaagtggctacctgaccccccaagcgcggtacatttgaaagcg 673
DB 1067 acggcatgtcagcgtaagtggctacctgaccccccaagcgcggtacatttgaaagcg 1126
QY 674 tgcagcaaaactggcccccgcgcgacaaaccccgacgacacaccccggtcatc 733
DB 1127 tgcagcaaaactggcccccgcgcgacaaaccccgacgacacaccccggtcatc 1186
QY 734 acaccaccccgatggcgccgcatcgcacccgacacccgacgacaccccggtcatc 793
DB 1187 acaccaccccgatggcgccgcatcgcacccgacacccgacgacaccccggtcatc 1246
QY 794 acagcgggctgctggcggggtgcgcgctgac 828
DB 1247 acagcgggctgctggcggggtgcgcgctgac 1281

RESULT 3
AAT33656
ID AAT33656 standard; DNA; 1535 BP.
XX AC AAT33656;
XX XX
XX DT 22-NOV-1996 (first entry)
XX XX
```

```
DE M. tuberculosis cellular uptake gene region.
XX KW Cellular uptake; cell entry; macrophage; passive immunisation;
XX KW vaccine; gene therapy; ds.
XX XX
XX OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
XX XX
XX FH Key Location/Qualifiers
XX CDS 181..810
XX FT /*tag= a
XX FT /label= ORF-1
XX FT /product= cell entry protein
XX FT 886..1535
XX FT /*tag= b
XX FT /label= ORF-2
XX FT /product= macrophage survival protein
XX XX
XX PN W09626275-A1.
XX XX
XX PD 29-AUG-1996.
XX XX
XX PF 20-FEB-1996; 96WO-US021155.
XX XX
XX PR 22-FEB-1995; 95US-0392210.
XX XX
XX PA (CORR ) CORNELL RES FOUND INC.
XX PI Riley LW;
XX XX
XX DR WPI; 1996-425086/42.
XX DR P-PSDB; AAW02301.
XX XX
XX PT DNA giving M. tuberculosis ability to enter mammalian cells -
XX PT and/or survive within macrophage(s), useful in vaccines to protect
XX PT mammals against Mycobacterium tuberculosis infection
XX XX
XX PS Claim 2; Page 41; 67pp; English.
XX XX
XX CC A DNA molecule (AAT33656) confers on Mycobacterium tuberculosis an
XX CC ability to enter mammalian cells and to survive within macrophages.
XX CC The encoded protein sequence is given in AAW02301. The DNA was obtd.
XX CC by ligating M. tuberculosis genomic DNA fragments into pBluescript II
XX CC vector and screening recombinant E. coli strains for Hela cell-
XX CC invasive clones. The DNA includes 2 separate coding regions (see
XX CC also AAT33657-58) coding for the cell entry (AAW02302) and macrophage
XX CC survival (AAW02403) proteins. It can be used to produce the cellular
XX CC uptake proteins used as vaccines or to facilitate uptake of other
XX CC materials, e.g., therapeutic genes.
XX SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 32.7%; Score 455; DB 17; Length 1535;
Best Local Similarity 100.0%; Pred. No. 4.1e-75;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 gatcgagagcgacgtcaaaagtattcgcccttttgcacactgcccgcgcg 433
DB 827 gatcgagagcgacgtcaaaagtattcgcccttttgcacactgcccgcgcg 886
QY 434 tggatgtgtccaccgagcgccgaaagcgacgtggtcagctggtacacccgacgcg 493
DB 887 tggatgtgtccaccgagcgccgaaagcgacgtggtcagctggtacacccgacgcg 946
QY 494 gtcccagacgagctggcccgctacgcccagcggtcatggactggctacacccgacgcg 553
DB 947 gtcccagacgagctggcccgctacgcccagcggtcatggactggctacacccgacgcg 1006
QY 554 acctaccacacacgacgacgcccgaacgacgtggtcagctggtacacccgacgcg 613
DB 1007 acctaccacacgacgacgcccgaacgacgtggtcagctggtacacccgacgcg 1066
QY 614 acggcatgtcagcgtaagtggctacctgaccccccaagcgcggtacatttgaaagcg 673
```


CC II vector and screening recombinant *E. coli* strains for HeLa cell-
CC invasive clones. It can be incorporated into a vector and used
CC for prodn. of recombinant macrophage survival protein, which is
CC useful in vaccines or for facilitating uptake of other materials,
CC e.g. therapeutic genes, into mammalian cells. A second DNA molecule
CC (AAT33657) codes for a protein (AAW02302) that confers on *M.*
CC tuberculosis the ability to enter mammalian cells.
XX
SQ Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

[illegible]

RESULT	6
AAV18649	
ID	AAV18649 standard; DNA; 650 BP.
XX	
XX	
AC	AAV18649;
XX	
DT	03-JUL-1998 (first entry)
XX	
DE	DNA for M. tuberculosis cellular uptake protein fragment.
XX	
KW	cellular uptake protein; vaccine; infection; ds.

XX	OS	Mycobacterium tuberculosis.
XX	XX	
XX	PH	Key
FT	FT	Location/Qualifiers
FT	FT	1..649
FT	FT	/*tag= a
FT	FT	/note= "stop codon not given"
XX	XX	
XX	PN	W09805784-A1.
XX	PN	
PD	XX	12-FEB-1998.
XX	XX	
PF	PF	06-AUG-1997; 97WO-US13056.
XX	XX	
XX	PR	07-AUG-1996; 96US-0689411.
XX	XX	
PA	PA	(CONN-) CONNAUGHT LAB LTD.
PA	PA	(CORR) CORNELL RES FOUND INC.

XX	Chong P, Riley LW;
PI	
XX	
DR	WPI: 1998-145620/13.
DR	P-PSDB; AAW47543.
XX	
XX	Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
PT	therapeutic agents e.g. antibiotics, also useful in vaccines
XX	
PS	Disclosure: Page 14; 82pp; English.
XX	
CC	The present sequence encodes a Mycobacterium tuberculosis cellular
CC	uptake protein fragment, which confers on M. tuberculosis an
CC	ability to survive within macrophages.
CC	The protein can be used in a vaccine to prevent M. tuberculosis
CC	infection, and provide for the uptake in cells of, e.g.
CC	antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
CC	raised against it can be used to treat mammals already exposed to
CC	M. tuberculosis, to induce a passive immunity and prevent disease
CC	occurrence.
XX	
SQ	Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;

Query Match	28.4%;	Score 396;	DB 19;	Length 650;
Best Local Similarity	100.0%;	Pred. No. 2.6e-64;		
Matches 396;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
Qy 433	gtgagtgtccaccgagcgccgagcgagcgacgactggcggcaagcgccgtcaatat	492		
Db 1	gtgagtgtgtccaccgagcgccgagcgacgactggcggcaagcgccgtcaatat	60		
Qy 493	cgteccgacgagctggccgctacgcacgaggtcatgtactgggtacacccgcagcg	552		
Db 61	cgteccgacgagctggccgctacgcacgaggtcatgtactgggtacacccgcagcg	120		
Qy 553	gaactcaacacacgaacgagcgccgcaaacgaggtcacctcgtgagcaaccagcaat	612		
Db 121	gaactcaacgacacgaacgagcgccgcaaacgaggtcacctcgtgagcaaccagcaat	180		
Qy 613	gacggcatgtcacggttaagtgtactgaaccccaagcgcgcgccaccttgaagc	672		
Db 181	gacggcatgtcacggttaagtgtactgaaccccaagcgcgcgccaccttgaagc	240		
Qy 673	gtctatgcgaactggcgcccccgaagatgaacccggaagacacaccccggtatc	732		
Db 241	gtctatgcgaactggcgcccccgaagatgaacccggaagacacaccccggtatc	300		
Qy 733	gaacacaccccggtatgcgcgcacatgaacccggaagacacaccccggtatc	792		
Db 301	gaacacaccccggtatgcgcgcacatgaacccggaagacacaccccggtatc	360		
Qy 793	cacgacgggtgtgtgcggggctgcgcgctgac	828		
Db 361	cacgacgggtgtgtgcggggctgcgcgctgac	396		

RESULT	7	
AAA92301		
ID	AAA92301	standard; DNA; 30690 Bp.
XX		
XX	AAA92301;	
AC		
XX		
DT	10-JAN-2001	(first entry)
XX		
XX		
DE	S. avermitilis	avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX		
KW	Streptomyces avermitilis;	avermectin aglycon synthase; biosynthesis;
KW	multifunctional enzyme;	polyketide; avermectin; veterinary drug;
KW	agrochemical;	ds.
XX		
OS	Streptomyces avermitilis.	
XX		

FH Key Location/Qualifiers
FT CDS 1..11919
FT /tag= a
FT /note= "Avermectin aglycon synthase protein"
FT CDS 11971..30690
FT /tag= b
FT /note= "Avermectin aglycon synthase protein"
XX WO2000050605-A1.
XX 31-AUG-2000.
XX 23-FEB-2000; 2000WO-JF01041.
XX 24-FEB-1999; 99JP-0046961.
XX (KITA) KITASATO INST.
XX Omura S, Ikeda H;
XX WPI: 2000-565458/52.
XX P-PSDB: AAB23749, AAB23750.
XX
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use -
XX
XX Claim 2; Page 66-134; 314pp; Japanese.
XX
XX The present sequence represents DNA which encodes avermectin aglycon
XX synthase proteins. Also described are: (1) polypeptides encoded by all
XX or part of the DNA; (2) expression vectors containing the DNA; (3) host
XX cells transformed by the vectors; (4) preparation of the polypeptides
XX by culture of the transformants; (5) preparation of avermectin aglycon
XX or its derivatives by culture of transformed avermectin-producing
XX microorganisms; and (6) oligonucleotides of 5-60 bases in length
XX containing sense or antisense sequences from the avermectin aglycon
XX synthase DNA. The enzymes are useful for the production of modified
XX forms of avermectin and of the intermediates in its biosynthesis, for
XX use as drugs, veterinary drugs and agrochemicals.
XX
XX Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

Db 21842 cactgtctcagacagcccgtagccgcccgcctcttcgcttccaggctccctcc 21901
Qy 641 tgaccccccaagcgcgccgacccctttgaagcgtgtgtagccaaactggcgcccccgcg 700
Db 21902 accgctctccaccgagcgtaccacatcacccccarLactacgocggacactctcgt 21961
Qy 701 cgaccaaccccgagcagccacccggtcatalgacaccccccggatcgccgcratcg 760
Db 21962 gggaaatccggcgcgccacctcgcggtatctctcctcaccgacgcccaccctca 22021
Qy 761 accggaracccgtagccaaagcrraagcraaccacgagcggtgctggcggtgcgcg 820
Db 22022 tcaccaacgcgcgacccctcatcaaaacatgcgcccgccaccatgacacccctca 22081
Qy 821 c-----gctgatcgctctcgtggaactcgtggccaaacaaacggtcttccggtctga 871
Db 22082 ccaccacacacacatcacccacgtctactgcccacgaaaaacgacctcgccatcgcc 22141
Qy 872 tctgtgtcatcacacccctgacagacgtgc-----aaacgcgcgcgcgaagggtctca 925
Db 22142 cctacaaacccctcctcctctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 22201
Qy 926 cggcgcgcgccacccctcctcctcctcctcctcctcctcctcctcctcctcctcctc 985
Db 22202 ccacctctgcccaraacaaaggcatcaaaacaaacacccctcccccacacacgcttcc 22261
Qy 986 actactcccccgcaagcggaggaglaccccagcggtatcttgcacacggccacacccctgg 1045
Db 22262 actccccccacacacacccctcctcctcctcctcctcctcctcctcctcctcctc 22321
Qy 1046 cgtgtgtatcacacacacccctcctcctcctcctcctcctcctcctcctcctcctc 1105
Db 22322 accaccccccacacacccctcctcctcctcctcctcctcctcctcctcctcctc 22381
Qy 1106 accgcggtgcacaaacgcgctgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1165
Db 22382 ccactactggacacaaagc 22441
Qy 1166 tcaccgcttggacgacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1222
Db 22442 tcaccaacacggcgtccaccacctcactcgaactcggacccgacacacccctccac 22501
Qy 1223 gccccgacacccgactcgcgcgaaagggctquaccacccacacacacaccccgccaca 1282
Db 22502 tcaccac 22561
Qy 1283 cctgactggtatctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1342
Db 22562 atccacccac 22621
Qy 1343 acccgaacgattcttccacacacacacacacacacacacacacacacacacacac 1386
Db 22622 accactac 22665

RESULT 8
AAZ17263
ID AAZ17263 standard; cDNA; 1337 BP.
AC AAZ17263;
XX
XX 12-OCT-1999 (first entry)
XX Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
XX detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX
XX WO9938972-A2.


```

XX PR 02-OCT-1998; 98US-0102748.
XX PR 11-MAR-1999; 99US-0123810.
XX PR 17-JUN-1999; 99US-0139650.
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX PT WPI; 2000-317716/27.
XX DR P-PSDB; AAY84727.
XX
XX PT New isolated polyketide synthase nucleic acid and polyketide compounds,
XX PT useful for treating e.g. transplant rejection, uveitis, multiple
XX PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
XX PT peripheral neuropathy.
XX
XX PS Example 1; Page 82-85; 126pp; English.
XX
XX CC The present sequence represents an NheII-XhoII fragment that encodes
XX CC module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX CC endogenous acyltransferase (AT) domain replaced by the AT domain of
XX CC module 12 of the rapamycin PKS (which is specific for malonyl CoA).
XX CC FK-506 is a potent immunosuppressant, and acts through initial formation
XX CC of an intermediate complex with protein immunophilins known as FK-506
XX CC binding proteins. The nucleic acids are used for producing polyketide
XX CC compounds. The polyketide compounds can be used as immunosuppressants to
XX CC prevent or treat transplant rejection, graft-versus-host disease or
XX CC uveitis. They can also be used for treating e.g. alopecia universalis,
XX CC autoimmune chronic active hepatitis, inflammatory bowel disease,
XX CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
XX CC also have neurotrophic activity and can be used to promote neurite
XX CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
XX CC and in intact animals, they promote regrowth of damaged facial and
XX CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
XX CC the brain. They can also be used for treating e.g. Parkinson's disease,
XX CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
XX CC peripheral neuropathies. They can also be used in agricultural and
XX CC veterinary applications.
XX
XX SQ Sequence 4466 BP; 634 A; 1796 C; 1433 G; 603 T; 0 other;

```

```

Query Match 6.3%; Score 87.4; DB 21; Length 4466;
Best Local Similarity 47.2%; Pred. No. 9.5e-08;
Matches 265; Conservative 0; Mismatches 296; Indels 0; Gaps 0;
QY 828 cgctccgggaactggccacaacacgggtcttcctcgtctcgtggtcaccacac 887
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3545 cgaccccgaggacataccaccccgccgcccacaccccgccaccccgctcgtcgcct 3604
QY 888 cctgacccgacctgcaaacccggcgccggaagggcttcacccggcgccacccgtctacc 947
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3605 gcaacaccacctcacacacccgacacacccctcgtctccacacccacccgacccgc 3664
QY 948 catggccggtgctcgcgatgaccagccagccacccactactctcccgcaagcggaag 1007
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3665 cggcgccacggctcaccgctcctcaccgacccgcccagacacccaccccgccatccg 3724
QY 1008 gtaccccccaggcgattctgaccacggcacacccctcgctgtatcacacaaacgct 1067
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3725 cctcatcgaaacggaccaccccccacacccctcctccctgcccacactcgccacccctga 3784
QY 1068 agcctcccgcccgagcgatcatgctgttgcgaacgacccggcggtgcacaaacccgg 1127
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3785 ccaccccccacctccgctcaccacacacccctcaccacccacccacccctccca 3844
QY 1128 ctgtgacccgacccgctaccacagccacccacacgctcaccgctggacgacccgg 1187
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
3845 caccaccccccacccacaccccccctcaaccccggaacacgcatcatcacccgg 3904
QY 1188 acgacccgacatcacccgagctgacctgctgctggtggccccgacacacgactcgccgaaa 1247
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 3905 cggctccggcaccctcgccggcgcatectctcccgccacccctggaaccccccacacctacct 3964
QY 1248 aggtctggaccaccccaacaacacacccacccgcccacacggaatggctaccacccccaacct 1307
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3965 cctctcccgaccccaaccccccgcgcaccccccgcacccacctcctcctcgacaculcgg 4024
QY 1308 cgacacggcgcaaccccgacacacacacacacacacacacacacacacacacacacacac 1367
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4025 cgaccccccacacacacacacacacacacacacacacacacacacacacacacacacac 14084
QY 1368 agacgacgacgacacacacacacacacacacacacacacacacacacacacacacac 1388
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4085 ccacacggcgccacacacacacacacacacacacacacacacacacacacacacacac 4105

RESULT 11
AAAL4661
ID AAA14661 standard; DNA; 4478 bp.
XX
AC AAA14661;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 3..4466
FT /*tag= a
FT /note= "no termination codon given"
XX
PN WO200020601-A2.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99US-0122886.
XX
PR 02-OCT-1998; 98US-0102748.
PR 11-MAR-1999; 99US-0123810.
PR 17-JUN-1999; 99US-0139650.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX
PT WPI; 2000-317716/27.
XX
DR P-PSDB; AAY84725.
XX
XX New isolated polyketide synthase nucleic acid and polyketide compounds,
XX PT useful for treating e.g. transplant rejection, uveitis, multiple
XX PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
XX PT peripheral neuropathy.
XX
XX PS Example 1; Page 76-79; 126pp; English.
XX
XX CC The present sequence represents an AvrII-XhoII fragment that encodes
XX CC module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX CC endogenous acyltransferase (AT) domain replaced by the AT domain of
XX CC module 12 of the rapamycin PKS (which is specific for malonyl units).
XX CC FK-506 is a potent immunosuppressant, and acts through initial formation
XX CC of an intermediate complex with protein immunophilins known as FK-506
XX CC binding proteins. The nucleic acids are used for producing polyketide

```

compounds. The polyketide compounds can be used as immunosuppressants to prevent or treat transplant rejection, graft-versus-host disease or uveitis. They can also be used for treating e.g. alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, or scleroderma. They also have neurotrophic activity and can be used to promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. They can also be used for treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or peripheral neuropathies. They can also be used in agricultural and veterinary applications.

Query Match	6.38;	Score 87.4;	DB 21;	Length 4478;
Best Local Similarity	47.2%;	Pred. No. 9.5e-08;		
Matches 265;	Conservative	0;	Mismatches 296;	Indels 0; Gaps 0;
QY	828	cgcctccgggaactggcccaacacacagcgtcttcctccgtctcgtatcgtggtgcaccacac	887	
Db				
Db	3557	cgaccctggagacataccaccctggccacacccgcccacacccgcccacccgctcgtgacgcct	3616	
QY	888	ctgtaccgactgtaaaacggccgcgcaagggtttcacggcgcgcgacccctgtacc	947	
Db				
Db	3617	gtaacaccacttcacacacccgacacacccctcatcgtccacacacacgaccccg	3676	
QY	948	catggccgagtgtacctgatgacagcacgcccacactactccccgcgaagcggg	1007	
Db				
Db	3677	cggcgccacgttcacccgctccaccgcacccgcccagagacacccaccgcatccg	3736	
QY	1008	gtaccctcaggcatcttcgacacggcacacccctggcctgtatcacacaaacgcct	1067	
Db				
Db	3737	ctctcat.aaacggacacccccacacccccctccctggcccaactcgcaacctcga	3796	
QY	1068	agctccccggcccgagcggtatcgtgttcgcacagccgcggctgcacaaacccg	1127	
Db				
Db	3797	ccacccccacttcgcctcacccacacacccctcacacccctcacctcacccccctca	3856	
QY	1128	ctgtgacgacgcgctaccacgacgaagcccacacgctcacccgctggaccagcacg	1187	
Db				
Db	3857	caccacacccacacacacacccccctcaacccggaacagcccatcatcatccg	3916	
QY	1188	acgcacccgacatcacggagctgacctggctgcggccgcgcacacacgacttccctgaaad	1247	
Db				
Db	3917	cggctccggacccctcgcggcatcctcgcgcgcactgaacacccccacacttact	3976	
QY	1248	aggctggacccaccacaaacacaccccgccacacggatggctacacacacccctcact	1307	
Db				
Db	3977	ctctcccgaccccccccgagccacccccggcaccacctcccttgcgacgcgg	4036	
QY	1308	cgacacggcccaacccccgcaccaacactttcacacccccgaacgattctctcacaa	1367	
Db				
Db	4037	cgacccccacacactgcgcacacccctcacccatcccccaacccccctaccgctctt	4096	
QY	1368	agacgacgacgacaaacccga	1388	
Db				
Db	4097	ccacacggcgccacccctcga	4117	

RESULT 12
AAAL14664
ID AAAL14664 standard; DNA; 4547 BP.
XX
AC
AC
AAAL14664;
XX
XX
08-AUG-2000 (first entry)
XX
XX
Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
DE
DE
FK-520: polyketide synthase; PKS; gene cluster; immunosuppressant;
KW
KW

KW	immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW	transplant rejection; graft-versus-host disease; alopecia universalis;
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW	neuropathic growth; nerve regrowth; Parkinson's disease;
KW	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW	peripheral neuropathy; ss.
XX	
XX	Synthetic.
OS	Streptomyces hygroscopicus.
XX	
XX	Key Location/Qualifiers
FT	CDS 9..4535
FT	/*tag= a
FT	/note= "no termination codon given"
XX	
XX	WC0200020601-A2.
XX	
XX	13-APR-2000.
XX	
XX	01-OCT-1999; 99WO-US22886.
XX	
XX	02-OCT-1998; 98US-0102748.
XX	
XX	11-MAR-1999; 99US-0123810.
XX	
XX	17-JUN-1999; 99US-0139650.
XX	
XX	(KOSA-) KOSAN BIOSCIENCES INC.
XX	
XX	Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX	
XX	WPI: 2000-317716/27.
XX	
XX	P-PSDB: AAY84728.
XX	
XX	New isolated polyketide synthase nucleic acid and polyketide compounds,
XX	useful for treating e.g. transplant rejection, uveitis, multiple
XX	sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
XX	peripheral neuropathy
XX	
XX	Example 1; Page 85-88; 126pp; English.
XX	
XX	The present sequence represents an NheII-XhoII fragment that encodes
XX	module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the
XX	endogenous acyltransferase (AT) domain replaced by the AT domain of
XX	module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
XX	FK-506 is a potent immunosuppressant, and acts through initial formation
XX	of an intermediate complex with protein immunophilins known as FK-506
XX	binding proteins. The nucleic acids are used for producing polyketide
XX	compounds. The polyketide compounds can be used as immunosuppressants to
XX	prevent or treat transplant rejection, graft-versus-host disease or
XX	uveitis. They can also be used for treating e.g. alopecia universalis,
XX	autoimmune chronic active hepatitis, inflammatory bowel disease,
XX	multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
XX	also have neurotrophic activity and can be used to promote neurite
XX	outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
XX	and in intact animals, they promote regrowth of damaged facial and
XX	sciatic nerves, and repair lesioned serotonin and dopamine neurons in
XX	the brain. They can also be used for treating e.g. Parkinson's disease,
XX	Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
XX	peripheral neuropathies. They can also be used in agricultural and
XX	veterinary applications.
XX	
XX	Sequence 4547 BP: 658 A; 1851 C; 1437 G; 601 T; 0 other;

	Query Match	6.3%	Score 87.4	DB 21	Length 4547
	Best Local Similarity	47.2%	Pred. No. 9.5e-08		
	Matches 265	Conservative	0	Mismatches 236	Indels 0
QY 828	cgctctcgggaacatcgggccaacacacacgggtctctcccgctctcgatctcgtggtgcaccacac	887			
Db 3626	cgaccccgagacatcacccacggcgccacacacggcgccacccggctctctgacgcgcct	3685			
QY 888	cctgaccgacctgcacacggcgccggcgcaagggtcttcacggcgcgcgacacctgtaur	947			


```
Db 1789527 GAGCACTACCGGTGAACCGCTAGCCCAACAGTTGACCGCACCGCCACCGCCCAACGCC 1789468
QY 367 agggctcatcgagagcgacgtcaaaagtattcgccctttttcgccacctacc 426
Db 1789467 AGGCCCTGATCGGAGGCGCAGTCAAGATGATTGCGCCCTTTTCGCCCACTTGCCC 1789408
QY 427 ggcggtggtgattgtccaccgcgcagcgccgcaagccgacacctggcgccgcaagcgct 486
Db 1789407 GCCGCGTGGATGTGTACACCGCCAGCGCGCGCGAAGCGACCTGGCGGCAAGCGCT 1789348
QY 487 caatatctccgacgagctggcccgctacgcccagcggtcatggtgactggctacacccc 546
Db 1789347 CAATATCTCTCCGACGAGCTGGCCCGCTAGCCCAAGCGCGCTATGGACTGGCTACACCC 1789288
QY 547 gacgagacctcaccgacacgaacgacgcccgaacgcaacgacgacacacccctgagcaaccg 606
Db 1789287 GACGGCGACCTACCGACACGAACGCGCGCGCGCAACCGCGACCAACCGCGACACCCG 1789228
QY 607 caatacgagcgatgtcacggctaaagtggctacgtgaccccccaagcgcgccacattt 666
Db 1789227 CAAATAGACGGCATGTACCGGTAAAGTGGCTACCTGACCCCGCAAGCGGGCCACTTT 1789168
QY 667 gaagccgtgttagcaaaactggccgcccccgcgacccaaacccccgacgacacaccccg 726
Db 1789167 GAAGCCGTGTAGCCAACTGGCCGCGCGCGCGCGCAACCGCGACCAACCGCGACACCCG 1789108
QY 727 gtcatcgacaccccccgatggcgccgcatcgacgcgcgacaccccgagccagcccaaa 786
Db 1789107 GTATCGACACACCCCGATGGCGCGCATCGACCGGACACCGCGACCGCAAGCCCAA 1789048
QY 787 cgcacacacgagcggtgtgtgcccgggtgctgcgcgctgacgctcctcgggaaactgggc 846
Db 1789047 CGCAACACACGAGGGGTGTGTGCGCGGCTGCGCGCGCTGATCGCTCGCGGAACTGGCG 1788988
QY 847 caacacacggtctccgctctcgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 906
Db 1788987 CACACAAACGGTCTTCCGCTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1788928
QY 907 ggcgcgcaagggcttcacggcgcgcgccctgtaccctgtaccctgtgacgacgacgacgac 966
Db 1788927 GCGCGCGCAAGGGCTTACCGCGCGCGCGCACCTGTCTACCATGGCCGATGTGTCTCCG 1788868
QY 967 atgaccagcagcccaactactccccgcaagcgggaggtaccctcccgagcgatcttc 1026
Db 1788867 ATGACCAAGCCACCCACCACTACTCCCGCAAGCGGAGGTACCCCGCAGCGATCTTC 1788808
QY 1027 gaccagggcacacctggcgctgtatcaccaaaacgctagcctcccgccgagcgg 1086
Db 1788807 GACCACGGCACACCTTGGCGCTGTATCACCAAAACGCTAGCCTCCCGCGCCAGCGG 1788748
QY 1087 atcatgtgttcgcaacgacgacgctgtgaccaaaccgctgtgacgacgacgacgac 1146
Db 1788747 ATCATGCTGTTCGCCAAACCGCGGCTGACCAAAACCGGCTGTGACGACCGGCTTAC 1788688
QY 1147 cacagcaagcccaactcagctgacgctgacgagcagcggaacgacgacgacgacgacgag 1206
Db 1788687 CACAGCAAGCCACCCACCACTACCGCTGGACGACGACGACGACGACGACGACGACGACG 1788628
QY 1207 ctgacctgacctgagcccgcaacacgactcgcgaaaaagctggaccaccccaaac 1266
Db 178627 CTGACCTGCTGCGGCGCGCGCGCAACCACTCGCGGAAAAAGGCTGGACCAACCAAC 1788568
QY 1267 aacacccacgcccacacggaatggctaccacacccccacacacgacgacgacgacgacgac 1326
Db 1788567 AACACCCAGGGCACACCGAATGGCTACCAACCCCACTCGACCAACCGGCAACCGTGG 1788508
QY 1327 accaacaccttcacaccccc 1347
Db 1788507 ACCTGTGAGATACACTACACC 1788487
```

RESULT 2

US-08-390-878-18

```
: Sequence 18, Application US/08390878
: Patent No. 5700683
: GENERAL INFORMATION:
: APPLICANT: Stover, Charles K.
: APPLICANT: Mahairas, Gregory G.
: TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: One Market Plaza, Steuart Street Tower, 20th
: STREET: Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM: disk
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/390,878
: FILING DATE: 17-FEB-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Hunter, Tom
: REGISTRATION NUMBER: 38,498
: REFERENCE/DOCKET NUMBER: 15371A-17
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/543/9600
: TELEFAX: 415/543/5043
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12412 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-390-878-18
```

Query Match 86.3%; Score 1202.4; DB 1; Length 12412;
Best Local Similarity 95.7%; Pred. No. 3.3e-246;
Matches 1290; Conservative 0; Mismatches 51; Indels 7; Gaps 5;

```
QY 7 cgggtagccgcgaacgaattgtcgaqgtcttggatcgctggatcgcgagctggaccgct 66
Db 84 CGAATAGCCGGGACGAGCTGTGCAAGTCTTTGATCGCTGGATGCGGACCTGACCGCT 143
QY 67 tggacgaggtgtcctttaggtattgacaccccaagacgctgctgctctggaacgctc 126
Db 144 TGGACGAGGTGCTTTTGGAGTGTACCCACCCCGAAGCGCTGCGGTCTCTGGAAACGTC 203
QY 127 tggatgtgttgctgcccggctaccgctgggtggtgacgctgtgtatcaaaccaacttgacg 186
Db 204 TGAATGCTTGGTGGCGCGCTACCGCGGTGCGGCGACACGTTGATCAACCAACTCGACA 263
QY 187 cccaagcgaaggaagaactggcgagcagctgtgtgctgctgctggcgaaccggcttac 246
Db 264 CCCAAGCAGGAGGAAGAACTGGCGCGCACGCTGTGCTGCGCTGGCGCAACCGGTTAC 323
QY 247 gcatcacaagcccgacgacgcccggcgcatcgcgacgcccgcgactctcgacacctcgt- 305
Db 324 GCATCACAAGCCGACGCGCGCTACGATCGCGACCGCGCGCGGATCTCGGACCTCTGTC 383
QY 306 cgagcactcacgggtgaaacgctagccccacag-ttgaccgcacacgcccacgcccacacg 364
Db 384 CGAGCACTCACCGCGCAACCGCTAGCCCGCACAGTTTGACCGCGCACCGCCACCGCCCAAGC 443
QY 365 ccagggcctgacggcg-agggcagcgtcaaaagtgtatcgccccctttttcg--cccacc 421
Db 444 CCAGGGCCTGTATCGCGGAAGGGCGACATCAAAAGTGTATTCGCGCCCTTTTTCGCGCCCAACC 503
```

QY	422	tgcgcgcgcggtgtagtgttc--accgcacaggccgcgaagcgcgacctggccgcgcaa	479
Db	504	TGCCCCCGCGGTGATGTGTCAAACCGCAGCGCGCGCGAAGCCGACCTGGCGCGCA	563
QY	480	agcgcgtcaatatcgtcccgacgagctggccggtctacgcccagcgaggcatggaactggcct	539
Db	564	ACCGTCAATATCGTCCCGACGAGCTTGGCCCGCTACGCCACGCGGTTCATCGACAGGT	623
QY	540	acaccccgacggcgacctcacgcacacgcgaacgcgcgcgcaaacgcggcatcaacctaaq	599
Db	624	ACACCCCGACGGCGACCTACCGGACACGAAACGCGCGCGCAACCGGGCATCACCTTGAG	683
QY	600	caacca--aatacacgagcatgtaacggttaagtggctacctgaaccccccaagcggcgc	659
Db	684	CAACACAGCAATACGACGGCATGTTCACGGCTTAAGTGGTACTCTGACCCCAAGCGCGGC	743
QY	660	cacctttgaagccgtgctagcctgaacaaactggccgcgcccgccgcgcgaacacccgcgacga	719
Db	744	CACCTTTGAAGCCGTGTAGCGAAACTGGCGCGCGCCCGCGGGGACCAACCGCGACGACCA	803
QY	720	caacccggtcatcgacacacaccccccgatgcggcgcccatcgacccgcgaacccgcgacga	779
Db	804	CACCCCGGTCTATCGACACCAACCCCGCATGGCGCGCCATCTGACCGCGGACACCCCGCAGCA	863
QY	780	agcccaacgcgaacacgcagcgggtgctggcggggtgcgcgctgtagctcctccggaaa	839
Db	864	AGCCCCACGGCAACACGACGGGCTGCTGGCGGGCTGCGCGGCTGATCGCTCTCGGGGA	923
QY	840	actgggccaacacaacagggtttcccgctctcgatctggttcacccacacctgaacgacct	899
Db	924	ACTGGGCCAACACAAACGGTCTTCCCGTCTCGATCTGTTGTTGTTACACACACCCCTCAGCGACCT	983
QY	900	gcacacggcgccgcgaaggcttcacggcgcgcgccacctgctaaccatggcgcgatg	959
Db	984	GCAAACCGGCGCGGCAAGGGCTTCACCGCGCGCGCACCTGTCTACCCATGGCCGATGT	1043
QY	960	gatccgatgacacgacgcccacacactactcccccgcaagcgagggtaccccccggc	1019
Db	1044	GATCCGATGACACGACGCGCCACACCTACTTCCCGCGCAAGCGGGAGGTACCCGCCAGGC	1103
QY	1020	gatcttcgacacgacacccccctggcgtgctatcacccaaacgcttagctctccccgc	1079
Db	1104	GATCTTGACACGGCACACCCCTGGCGGCTGTATCACACCAACGCCCTAGCTTCCCCCGC	1163
QY	1080	ccagcgatcatcgttctgcgaacgacgcggcgtgcacccaaaccccgctgtgtacgacc	1139
Db	1164	CCAGCGGATCATGCTGTTCGCAACACGCGCGGTGTCACCAACCCGGCTGTGACACAC	1223
QY	1140	ggctacacagcgaagcccacacgctcacccgctggactagcgcgcgcgcgcacat	1199
Db	1224	GGCTTACACAGCAAGGCCACACGTCACCGGCTGGACCAACACCGGACGACACCGACAT	1283
QY	1200	cacggagctgacctggcctgcgccccgcgaacacgcgactgcgcgaataagcctggaccac	1259
Db	1284	CACCGACCTTACCCCTGGCCTGGACCGCCGACACCGAGCTGCCCGAAAAGGCTGGACCAC	1343
QY	1260	ccacaacaaccccacggccacacccgaatgctaccaccacccacctgcacacgacga	1319
Db	1344	CCGCAAAACACCCACGGCCACACCGATGGCTACCAACACCCACCTCGACCACGGCCA	1403
QY	1320	accccgacacaacaccttcaccacccc	1347
Db	1404	ACCGTGGACCTGTGAGATACACTACCC	1431

RESULT 3
 US-08-464-052-1
 ; Sequence 1, Application US/08464052
 ; Patent No. 6008201
 ; GENERAL INFORMATION:
 ; APPLICANT: Riley M.D., Lee W.
 ; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
 ; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,052
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-464-052-1

[illegible]

RESULT 4
US-08-461-002-1

```

; Sequence 1, Application US/08461002
; Patent No. 6214543
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,002
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

Query Match 32.7%; Score 455; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 3.9e-88;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 gatcgagcgagcgacgtcaagtgtattcgcccttttgcgccacctgcccgcgcgc 433
DB 827 GATCGGCGAGGCGCAGCTCAAGTGAATTCGCGCCCTTTTCGCCACCTGCGCGCGCG 886

QY 434 tggatgtgtccaccgcccagccgacgacgtggcgacgtggcgacgtggcgacgtg 493
DB 887 TGGATGTGTCCACCCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946

QY 494 gtcccgacgagctggcccgctacgcccaggggtcatggactggctacaccccgagcgc 553
DB 947 GTCCCGACGAGCTGGCCCGCTACGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006

QY 554 acctcaccgacacccaagcgccgcaaacgcggcatcacctgagcaacacgaataac 613
DB 1007 ACCTCACCACACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066

QY 614 acggcatgtcacggctaaagtgtacctgaccccccagcgcgccacatttgaagcgc 673
DB 1067 ACGGCATGTACGGCTAAGTGGCTACCTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1126

QY 674 tctagcacaactggccgcccggcgagcacaaccccccgagcacaaccccggtcatcg 733
DB 1127 TGCTAGCACAACCTGGCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186

QY 734 acacacccccgatggcgcccgccatgacgcgacaccccgagcgaagcccaacgaacc 793
DB 1187 ACACACCCCGGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1246

QY 794 acgacgggctgtggtggcggtgctgctgctgctgctgctgctgctgctgctgctgct 828
```

```

DB 1247 ACGACGGGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281

RESULT 5
US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pele
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,411
; FILING DATE:
; CLASSIFICATION: 516
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/187
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-689-411-1

Query Match 32.7%; Score 455; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 3.9e-88;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 gatcgagcgagcgacgtcaagtgtattcgcccttttgcgccacctgcccgcgcgc 433
DB 827 GATCGGCGAGGCGCAGCTCAAGTGAATTCGCGCCCTTTTCGCCACCTGCGCGCGCG 886

QY 434 tggatgtgtccaccgcccagccgacgacgtggcgacgtggcgacgtggcgacgtg 493
DB 887 TGGATGTGTCCACCCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946

QY 494 gtcccgacgagctggcccgctacgcccaggggtcatggactggctacaccccgagcgc 553
DB 947 GTCCCGACGAGCTGGCCCGCTACGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006

QY 554 acctcaccgacacccaagcgccgcaaacgcggcatcacctgagcaacacgaataac 613
DB 1007 ACCTCACCACACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066

QY 614 acggcatgtcacggctaaagtgtacctgaccccccagcgcgccacatttgaagcgc 673
DB 1067 ACGGCATGTACGGCTAAGTGGCTACCTGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1126

QY 674 tctagcacaactggcgcccggcgagcacaaccccccgagcacaaccccggtcatcg 733
```

```
Db 1127 TGCTAGCAAACTGGCGCGCCCGCGGCGCAACCAACCCGAGACACACACACCCCGGTATCG 1186
QY 734 acaccacccccgatggcgcccgccatcgaccgcygacacccgagccaaagcccaacgcaacc 793
Db 1187 ACACCAACCCCGATGGCGCGCGCATGACGACCGGAGACACCGCAAGCCCAACGCAACC 1246
QY 794 acacaggggtgctggcggggtgcgcgcgctgac 828
Db 1247 ACACGGGCTGCTGGCGGCGCTGGCGGCGCTGATC 1281

RESULT 6
PCT-US94-09863-1
; Sequence 1, Application PC/TUS9409863
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR
; UPTAKE OF MYCOBACTERIUM TUBERCULOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Michael L. Goldman
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09863
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman Mr., Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/180 (D-1485)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1000
; TELEFAX: (716)-263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
PCT-US94-09863-1
```

```
Query Match 32.7%; Score 455; DB 5; Length 1535;
Best Local Similarity 100.0%; Pred. No. 3.9e-88;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 374 gatcggcgagcgacacgaatgattcgcccttttgcacacccctgcgcgcgcgg 433
Db 827 GATCGGCGAGCGCGACGTCAAGATGATTCGCGCCCTTTTCGCCACCTGCCCGCGCGG 886
QY 434 tggatgltccaccgcgagccgcaagccgacccgagcgaagcgaagcgcgctcaatc 493
Db 887 TGGATGTGTCCACCCCGCAGCGCCGCGGACCTGGCGGCAAAAGCGGTCAATATC 946
QY 494 gtcccgacgagctggcccgctacgcccagcggtcgtgactggtcagcaccgcgagcg 553
Db 947 GTCCCGACGAGCTGGCGCGCTACGCCACGCGGTATGGACTGGCTACACCCCGCGGG 1006
QY 554 acctcaccgacacgcgagcgcccgcaaacgcggtatcaccttgagcaaccgcaatag 613
Db 1007 ACCTCACCACACCGAAGCGCGCCGCAAAAGCGGCGATCACCTTGAGCAACCAATACG 1066
```

```
QY 614 acggcagtgacggctaaagtggctacctgaccccccaagcgcgccacctttgaagccg 673
Db 1067 ACGGCATGTACGGCTAAGTGGCTACCTGACGCCCCCAAGCGCGGCACCTTTGAAGCCG 1126
QY 674 tgctagcacaactgctcgcccccgctgagcccaacccccgagcaccacccccggtcacy 733
Db 1127 TGCTAGCAAACTGGCGCGCGCGCATGACGACCGGAGACACCGCAAGCCCAACGCAACC 1186
QY 734 acaccacccccgatggcggggtgcgcgcgctgac 828
Db 1187 ACACGGGCTGCTGGCGGCGCTGGCGGCGCTGATC 1281
QY 794 acgacggggtgctggcggggtgcgcgcgctgac 828
Db 1247 ACGACGGGCTGCTGGCGGCGCTGGCGGCGCTGATC 1281
```

```
RESULT 7
US-08-464-052-5
; Sequence 5, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263 1304
; TELEFAX: (716) 263 1600
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-5
```

```
Query Match 28.4%; Score 396; DB 3; Length 650;
Best Local Similarity 100.0%; Pred. No. 1e-75;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 433 gtggatgtgtccaccgcgagccgagccgagcgtgagcgtgagcgtgagcgtgagcgt 492
Db 1 GTGGATGTGTCCACCCCGCAGCGCCGCGGAGCCGACCTGGCGGCAAAAGCGGTCAATAT 60
QY 493 ggtccgacgagctggcccgctacgcccagcggtcgtgactggtcagcaccgcgagcgc 552
Db 61 GGTCCGACGAGCTGGCGCGCTACGCCACGCGGTGATGGACTGGCTACACCCCGACGC 120
QY 553 gacctcaccgacacgcgagcgccgcaaacgcggcgtcacctgagcaaccgcaatcac 612
```


Db 3870736 ccgtggacctgtgagatacactacacc 3870762

RESULT 11

```

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLETSCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103-840A-1
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match 7.3%; Score 101.2; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 1393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          5.6%; Score 78.4; DB 4; Length 4403765;
Best Local Similarity 97.2%; Pred. No. 8.9e-08;
Matches 1344; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 7 cgggtagccgcgaacgattgttcgaggtttttagctgctggtatgcgagctggacgcgt 66
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780705 CGAATAGCCGGAGGAGCTTGTGAGGTCTTTGATGCGCTGGATGCCGAGCTGGACCGCT 1780646

QY 67 tgaacaggtgtcttttgaggtgttgacacaccccaagaaacggtgctcttggaacgtc 126
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780645 TGAACGAGGTGTCTTTTGAGGTGTGTGACACCCCGGAACGGCTGCGGTCTGTGAAACGTC 1780586

QY 127 tgaatgtcttggtgcgcgggtaccggcggttggttcacgctgtgatcaacacacttgacg 186
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780585 TGAATGTCTTGTGCGCGGTACCGCGCTGGCGCACACGCTTGATCAACCAACTCGACA 1780526

QY 187 ccaacgcagcaggaagaactggcggaactgtgctgctgctgctgctgctgctgctgcttac 246
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780525 CCCAAGCCAGCAGGAAGAACTGGGCGCACGCTGTGCTGCGCGCTGGCCAAACCGTTAC 1780466

QY 247 gcatcaccagccgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780465 GCATCAACCAAGCCAGCAGCGCCCTACGCATCGCGGACGCGCGCGCTGTGACCTCGTC 1780406

QY 307 gacactaccggtgaaccgctagccccacagttgacgcgcacccgcgcgcgcgcgcgcgcgcgc 366
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780405 GAGCACTCACCGGGAACCGCTAGCCGCCACAGTTGACCGCCACCGCCACCGCCCAACGCC 1780346

QY 367 agggcctgacgcgagcgcagctcaagtgattcgccctttttggccacactgccc 426
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780345 AGGGCTGATCGGCGAGCGGCACATCAAGATGATTCGGCGCCCTTTTTCGCCACCTGCC 1780286

QY 427 gccgcggtggtgtgtccacccgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 486
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780285 GCGCGGTGGATGTGTCCACCCGCGCAGCGCGCGGAAACCGACCTGGCGCGCAAGCGCCT 1780226

QY 487 caatatgtccgcagagctggccgctacgccagcgggtataggactgctacacccc 546
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780225 CAATATCTCCGACGAGCTGGCCCGCTACGCCACGGGTCTATGGACTGGCTTACACTTC 1780166

QY 547 gacgcgactcacgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 606
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780165 GACGGCGACCTCACGACACCGACCGCGCGCCGCAAAAGCGCGCATCACCTGACCAACCCAG 1780106

QY 607 caatacgcgcgcgtacgcgctgaagtgtctactgaccccccaagcgcgcgcgcaccttt 666
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780105 CAATACGACCGCATGTACAGGCTAAAGTGGCTAGCTGACCCCGCAAGCGCGGCACCTTT 1780046

QY 667 gaagcgtgtgtagcaaaactggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 726
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1780045 GAACCGTGTAGCCAAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779986

QY 727 gtcatcgacacaccccgatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 846
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779985 GTCATCGACACCAACCCCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1779926

QY 787 cgaacacacgcgcgtactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 846
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779925 CGCAACCAACGACGGGTGTGTGGCGCGGGCTGCGCGCGCTGATCGCTCCGGGAAACTGGGC 1779866

QY 847 caacacacggtcttcccgctctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 906
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779865 CAACACAAGGTTCTCCGCTCTCGATCGTGGTCAACCAACCAACCGCTGACCGACCTGCAAAAC 1779806
```

```
QY 907 ggccgcgcgaaggccttcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 966
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779805 GGCGCGGCAAGGGCTTCACCGCGCGCGCGCACCTGCTTACCATATGGCGATGTGATCCGC 1779746

QY 967 atgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1026
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779745 ATGACGACGACGCGCCACCACTACTCCCGCCCAAGCGGAGGTACCCCGACGGCATCTTC 1779686

QY 1027 gaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1086
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779685 GACCACGCGCACATCCCTGGCGCTGTATACACCAAAAGCTTAGCTCCCGGCGCCAGCGG 1779626

QY 1087 atcatgtgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1146
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779625 ATCATGTGTGTGCGCAACGACCGCGGCTTGACCAAAATCCGGCTGTGACGACCGCGCTAC 1779566

QY 1147 cacagccaaagccacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1206
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779565 CACAGCCAAGCCCAACCACTACCGCGCTGGACGACCGGACCGACGACATCACCGAC 1779506

QY 1207 ctgacccctggcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1266
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779505 CTCACCCCTGGCGTGCAGACCCCGCGACACCGACTCGCCCAAAAGGCTGGACACCCACAAA 1779446

QY 1267 aacacccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1326
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779445 AACACCCACGCGCACACCGCAATGCTTACCAACACCCCGCTGACGACCGGCAACCGCGC 1779386

QY 1327 accaaacacttcacaccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1386
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1779385 ACCAAACACTTCCACCAACCGACGAGAGCTGCTACGCCACAACGAGGACAACCAACGAC 1779326

QY 1387 gat 1389
|||
Db 1779325 GAT 1779323

RESULT 13
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/11903451
; CURRENT APPLICATION NUMBER: US/09/140,114
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match          5.1%; Score 70.8; DB 2; Length 1931;
Best Local Similarity 43.1%; Pred. No. 8.9e-07;
Matches 342; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

QY 578 gcaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 258 gccacctgtctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtc 317

QY 638 acctgaccccccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 697
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 318 tctctcccgctcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgtcctcgt 377

QY 698 gcgcgaccacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 757
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
```

[illegible]

```

RESULT      14
US-08-858-003-1
: Sequence 1, Application US/08858003
: Patent No. 6060234
: GENERAL INFORMATION:
: APPLICANT: Katz, Leonard
: APPLICANT: Scassi, Diane L.
: APPLICANT: Summers Jr., Richard
: APPLICANT: Ruan, Xiaohan
: APPLICANT: Pereda-Lopez, Ana
: APPLICANT: Kakavas, Stephan J.
: TITLE OF INVENTION: NOVEL POLYK
: TITLE OF INVENTION: AND RECOMBI
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Abbott Laboratories
: STREET: 100 Abbott Park Rd.
: CITY: Abbott Park
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 2.0

```

1 CURRENT APPLICATION DATA:
 2 APPLICATION NUMBER: US/08/858,003
 3 FILING DATE: 16-MAY-1979
 4 CLASSIFICATION: 435
 5 PRIORITY APPLICATION DATA:
 6 APPLICATION NUMBER:
 7 FILING DATE:
 8 ATTORNEY/AGENT INFORMATION:
 9 NAME: Dianne Casuto
 10 REGISTRATION NUMBER: P-40,943
 11 REFERENCE/DOCKET NUMBER: 4952.US.P2
 12 TELECOMMUNICATION INFORMATION:
 13 TELEPHONE: (847)-938-3137
 14 TELEFAX: (847)-938-2623
 15 TELEX:
 16 INFORMATION FOR SEQ. ID. NO.: 1:
 17 SEQUENCE CHARACTERISTICS:
 18 LENGTH: 925 base pairs
 19 TYPE: nucleic acid
 20 STRANDEDNESS: double
 21 TOPOLOGY: linear
 22 US-08-858-003-1

[illegible]

RESULT 15
US-09-078-166-1
; Sequence 1, Application US/09078166
; Patent No. 6063561
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 18:47:35 ; Search time 3835.7 Seconds
(without alignments)
3902.513 Million cell updates/sec

Title: US-09-785-904-1

Perfect score: 1393

Sequence: 1 tgggttcggtagccgcgaa.....cgacgacaacccgattgac 1393

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_esttr:*

4: em_estom:*

5: em_estpi:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_hic:*

10: gb_estl:*

11: gb_est2:*

12: gb_hic:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_fod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	105.2	7.6	1197	11	BI416470 hasp001xa
C 2	104.6	7.5	1223	11	BI416573 hasp001xg
C 3	100.4	7.2	1015	11	BI416650 hasp001xm
C 4	99.2	7.1	924	10	BE455162 HVSMEb009
C 5	98.4	7.1	1224	11	BI416537 hasp001xe
C 6	98	7.0	931	11	BI416575 hasp001xg
C 7	97.2	7.0	1159	11	BF260036 HVSMEf002
C 8	96.6	6.9	289	10	AI864419 w152f10.x
C 9	95.8	6.9	863	11	BI416699 hasp001xp
C 10	95.8	6.9	1348	11	BF628951 HVSMEb000
C 11	95.6	6.9	1016	13	AL249930 Tetraodon
C 12	95.4	6.8	825	11	BG786255 SEAMC006

C 13	95	6.8	918	13	A0895329
C 14	95	6.8	966	13	AQ748706
C 15	94.4	6.8	782	11	BG809582
C 16	94.4	6.8	855	10	AL572700
C 17	94.4	6.8	1006	10	BE034822
C 18	94.4	6.8	1035	11	BF254665
C 19	94	6.7	873	11	BG786249
C 20	94	6.7	1192	11	BF256672
C 21	93.6	6.7	897	11	BI416517
C 22	93.6	6.7	1299	11	BF267681
C 23	93.2	6.7	1050	11	BF267685
C 24	93	6.7	952	11	BG441341
C 25	92.4	6.6	1088	11	BF256580
C 26	92	6.6	1195	11	BF256755
C 27	91.4	6.6	1039	11	BF259783
C 28	90.2	6.5	1042	11	BI416535
C 29	90	6.5	1280	11	BF864339
C 30	89.6	6.4	1166	11	BF256751
C 31	89.2	6.4	1186	11	BF256757
C 32	89	6.4	1325	11	BF256543
C 33	88.4	6.3	1122	5	BF628976
C 34	88.2	6.3	884	13	AQ782679
C 35	88	6.3	883	13	AZ186646
C 36	88	6.3	888	13	AZ189327
C 37	87.6	6.3	915	13	AZ195437
C 38	87.4	6.3	1086	11	BG809658
C 39	87.2	6.3	1080	11	BI416623
C 40	87	6.2	834	11	BI416657
C 41	87	6.2	918	11	BG809598
C 42	87	6.2	1504	10	BE421615
C 43	86.8	6.2	806	11	BI416659
C 44	86	6.2	1277	11	BG447302
C 45	85.8	6.2	893	13	AZ199082

ALIGNMENTS

RESULT 1

BI416470/c

LOCUS

DEFINITION

BI416470

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BI416470 1197 bp mRNA EST
hasp001xa01f Heterobasidion annosum - Scots pine infection stage
subtraction cDNA library (hasp) Heterobasidion annosum/Pinus
sylvestris mixed EST library cDNA clone hasp001xa01f, mRNA
sequence.
BI416470
BI416470.1 GI:15187493
EST.
Heterobasidion annosum/Pinus sylvestris mixed EST library.
Eukaryota; mixed EST libraries.
1 (bases 1 to 1197)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
Location/Qualifiers
1..1197
/organism="Heterobasidion annosum/Pinus sylvestris mixed
EST library"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"

```
/dev_stage="Seedling roots of scots pine were infected for
6 weeks with H. annosum"
notes="Vector: pT-Adv; Site:1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."

```

[illegible][illegible]

```

RESULT      2
BI416573/c
LOCUS
DEFINITION
            1223 bp      mRNA      EST      15-AUG-2001
            hasp001xglfr Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Heterobasidion annosum/ptinus
            sylvestris mixed EST library cDNA clone hasp001xglfr, mRNA
            sequence.
            BI416573
            BI416573.1  GI:15187596
            EST.
            Heterobasidion annosum/ptinus sylvestris mixed EST library.
            Heterobasidion annosum/ptinus sylvestris mixed EST library.
            Eukaryota; mixed EST libraries.
            1 (bases 1 to 1223)
            Asiegbu, F.O., Nahai-Kova, J., Choi, W., Stenlid, J. and Dean, R.A.
            Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
            unpublished (2001)
            Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
            Sweden.
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mvkpat.slu.se

```

FEATURES	source	seq. primer: 17 primer.
location/Qualifiers		
1. .1223		
Organism	"Heterobasidion annosum/pinus sylvestris mixed EST library"	
Adb_xref	"taxon:169015"	
Accession	"hsp001x11f"	
Accession_lib	"heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"	
/dev_stage	"Seedling roots of scots pine were infected for 6 days with H. annosum"	
/note	"vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (RP5)."	
BASE COUNT	49 a 24 c 1047 g 34 t 69 others	
ORIGIN		

[illegible]

[illegible]

RESULT 3
BI416650/C

LOCUS	B1416650	1015 bp	mRNA	EST	15-AUG-2001
DEFINITION	hsp001xm01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Heterobasidion annosum/Pinus sylvestris mixed EST library cDNA clone hasp001xm01f, mRNA sequence.				
ACCESSION	B1416650				
VERSION	B1416650.1	GI:15187673			
KEYWORDS	EST.				
SOURCE	Heterobasidion annosum/Pinus sylvestris mixed EST library. Heterobasidion annosum/Pinus sylvestris mixed EST library. Eukaryotes: mixed EST libraries.				
ORGANISM	1 (bases 1 to 1015)				
REFERENCE	Asiegbu, F.O., Nahaiikova, J., Choi, W., Stenlid, J. and Dean, R.A.				
AUTHORS	Expressed sequence tags of randomly selected cDNA clones from the				
TITLE	interaction of the root rot fungus (Heterobasidion annosum) with				
JOURNAL	seedling roots of Scots pine (Pinus sylvestris) with				
COMMENT	Unpublished (2001)				
	Contact: Fred O. Asiegbu				
	Dept. of Forest Mycology & Pathology				
	Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,				
	Sweden				
	Tel: +46 18 67 15 98				
	Fax: +46 18 30 92 45				
	Email: Fred.Asiegbu@mykopat.slu.se				
	Seq primer: '7 primer.				
FEATURES	location/Qualifiers				
source	1..1015				
	/organism="Heterobasidion annosum/Pinus sylvestris mixed				
	EST library"				
	/db_xref="taxon:169015"				
	/clone_lib="hasp001xm01f"				
	/clone_lib="Heterobasidion annosum - Scots pine infection				
	stage subtraction cDNA library (hasp)"				
	/dev_stage="Seedling roots of Scots pine were infected				
	6 days with H. annosum"				
	/note="Vector: pr-Adv; Site_1: EcoRI; The subtractive				
	hybridization cDNA library was constructed from Scots				
	pine roots infected for 6-days with mycelia of				
	Heterobasidion annosum (FP5)." 16 t.				
BASE COUNT	21 a 6 c 961 g 16 t.				
ORIGIN					

[illegible]

JOURNAL					
seedling roots of Scots pine (<i>Pinus sylvestris</i>)					
Unpublished (2001)					
CONTACT					
Contact: Fred O. Asiegbu					
Dept. of Forest Mycology & Pathology					
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,					
Sweden					
Tel.: +46 18 67 15 98					
Fax: +46 18 30 92 45					
Email: Fred.Asiegbu@mykopat.slu.se					
Seq primer: T7 primer.					
Location/Qualifiers					
source					
1. .931					
/organism="Heterobasidion annosum/Pinus sylvestris mixed					
EST library"					
/db_xref="taxon:169015"					
/clone_lib="Heterobasidion annosum - Scots pine infection					
stage subtraction cDNA library (hasp)"					
/dev_stage="Seedling roots of scots pine were infected for					
6 days with H. annosum"					
/note="Vector: pT-Adv; Site_1: ECORI; The subtractive					
hybridization cDNA library was constructed from scots					
pine roots infected for 6-days with mycelia of					
Heterobasidion annosum (FP5)."					
BASE COUNT 9 a 7 c 884 g 15 t 16 others					
ORIGIN					
Query Match 7.0%; Score 98; DB 11; Length 931;					
Best Local Similarity 44.2%; Pred. No. 5.8e-08;					
Matches 365; Conservative 0; Mismatches 461; Indels 0; Gaps 0;					
QY	541	cacccgcagcgacctcaactgacacggaaaggccccgaacggcgcatcaccttgagc	600		
DB	877	CCCCCCCCCCCCCCCCCCCCCCCCCCCNCGCCCCCGCCGCGCCGCGCCGCGC	818		
QY	601	aaccgacaataacgagcgatgtacgggttaagtggctacctgaccccccaagcgggcc	660		
DB	817	CACCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	758		
QY	661	acctttgaagcgtgttagcgaacactggcgccgccggcggaaccaaacccgacgaccac	720		
DB	757	CCCCCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	698		
QY	721	acccggtgatcgacacaccccgcgatggcgccgcatatgacccgcgacacgcgagcraa	780		
DB	697	CCCCCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	638		
QY	781	gcceaacgaaccacgagcggtgctggccgggctggcgcgctgatcgcttccttqaaa	840		
DB	637	CGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	578		
QY	841	ctgggcgaacacaaagcgtcttcctgctctgatctgttgtaccacacacctgaccgacctg	900		
DB	577	CCACCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	518		
QY	901	caaacgcgcccgcggaagggttacggcgcgcgacacctctaccatgacccgatgtg	960		
DB	517	CGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	458		
QY	961	atccgatgaccagcacgcccacactactctcccccgaagcggagggtaccccccaggcg	1020		
DB	457	CGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	398		
QY	1021	attcttgaccagcgcaaccccttgcgctgtatatcaacaaacgcttagctctcccgccc	1080		
DB	397	CGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGC	338		
QY	1081	cagcgatcatctgttgccaacgacgcggtgtcaccaaaccccggtgtgtgacgcacgc	1140		
DB	337	CNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCC	278		
QY	1141	gctaccacagcgaagccaccacgtcacgcgctggacgacgcggaacgcacccgacatc	1200		

```
QY 556 ctccaccgacaccgaacgcccgcgcaaacgcgcatcacccctgagcaaacaccgaatacagac 615
Db 266 TCCTCCCTCCACACACCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCG 325
QY 616 ggcattgtcacgggtaagtgtactactgacaccccccaagcgcgccacacacacacacacac 675
Db 326 ACCCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCG 385
QY 676 ctacgcaaacctggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 735
Db 386 CCNNCCCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 445
QY 736 accaccccccgtgagcgccatcgacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 795
Db 446 NCNCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 503
QY 796 gacgggctgtggcggggctgcgcgctgacgctcctccggggaactgggccaacacacacac 855
Db 504 NCCCNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNN 563
QY 856 ggtcttccctgctcgtatgctgtgtaccacacacacacacacacacacacacacacacacacac 915
Db 564 ACCCATCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 623
QY 916 aagggttcacgcggcgggcgacccctgctaccacacacacacacacacacacacacacacacac 975
Db 624 TCACCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCG 683
QY 976 cagcgcacacactactcctcccgcaagcgagggtaccctcccgagcgatcttcgacacacacac 1035
Db 684 CCCCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 743
QY 1036 acacccctggcgctgtatcaacacacacacacacacacacacacacacacacacacacacac 1095
Db 744 CCACCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCG 803
QY 1096 ttcgccaacgacgcgctgacacacacacacacacacacacacacacacacacacacacacac 1155
Db 804 CAATCTCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 863
QY 1156 gccacacacgctacgcgctggacgacgacgacgacgacgacgacgacgacgacgacgacgac 1215
Db 864 CCCCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 923
QY 1216 gcttgcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1275
Db 924 NCCCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 983
QY 1276 ggcacacacggaatggttacacacacacacacacacacacacacacacacacacacacacac 1335
Db 984 CCNANNCCCTCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 1043
QY 1336 ttcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1366
Db 1044 CCACACACACCTTACACACACACACACACACACACACACACACACACACACACACACACAC 1074

RESULT 8
AI864419/c 289 bp mRNA EST 07-MAR-2000
LOCUS w152f10.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2428555 3'
DEFINITION similar to TR:033201 033201 HYPOTHETICAL 49.4 KD PROTEIN. ;, mRNA
sequence.
ACCESSION AI864419
VERSION AI864419.1 GI:5528526
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 289)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

TITLE
JOURNAL
COMMENT

National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
www-bio.lnl.gov/hbrp/image/image.html
Insert length: 621 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 282.

FEATURES
source

1..289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2428555"
/tissue_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCACCAATCTGAAGTGGAGCGCCGATAGGTTTTTTTTTTTTTTTTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 40 a 106 c 111 g 32 t
ORIGIN

Query Match 6.9%; Score 96.6; DB 10; Length 289;
Best Local Similarity 59.1%; Pred. No. 9.3e-08;
Matches 165; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 80 ttttgagggtgttgaccac 139
Db 279 TTTGACGCGGTTAGTCCCG 111
QY 140 ggc 199
Db 219 CCAGCGGTTCGAGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCG 160
QY 200 ggaagaaatagggc 259
Db 159 GCGCGAGTTGCGTGGCG 100
QY 260 cgaaccccccgc 319
Db 99 CGAGCG 40
QY 320 tgaacgc 358
Db 39 CGAGCGCGAGTTCG 1

RESULT 9
BI416699/c
LOCUS BI416699
DEFINITION hasp001xp01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Heterobasidion annosum/Finus sylvestris mixed EST library cDNA clone hasp001xp01f, mRNA sequence.
ACCESSION BI416699
VERSION BI416699.1 GI:15187722

BI416699 863 bp mRNA EST 15-AUG-2001
hasp001xp01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Heterobasidion annosum/Finus sylvestris mixed EST library cDNA clone hasp001xp01f, mRNA sequence.

KEYWORDS	EST.
SOURCE	Heterobasidion annosum/Pinus sylvestris mixed EST library. Heterobasidion annosum/Pinus sylvestris mixed EST library Eukaryota; mixed EST libraries.
REFERENCE	1 (bases 1 to 863)
AUTHORS	Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE	Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL	Unpublished (2001)
COMMENT	Contact: Fred O. Asiegbu Dept. of Forest Mycology & Pathology Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden Tel.: +46 18 67 15 98 Fax: +46 18 30 92 45 Email: Fred.Asiegbu@mykopat.slu.se Seq primer: T7 primer.
FEATURES	Location/Qualifiers
source	1..863
	/organism="Heterobasidion annosum/Pinus sylvestris mixed EST library" /db_xref="taxon:169015" /clone="hasp001xp01f" /clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)" /dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum" /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FP5)."
BASE COUNT	12 a 21-c 700 g 12 t 118 others
ORIGIN	
Query Match	6.9%; Score 95.8; DB 11; Length 863;
Best Local Similarity	43.1%; Pred. No. 1.4e-07;
Matches 292; Conservative	0; Mismatches 386; Indels 0; Gaps 0;
QY	cgccccgcgagcaaacccggacgacacaccgccgtcatcgacacaccccgcgatg 748
Db	
B38	CCCCCCCCCCCCCCCCNCCCNCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCNN 779
QY	cggcgccatgacgcgcgacacccgcgcgaaggcccaacgcaacgacgagcgctgtg 808
Db	
B78	CCCCNNCCCCCCCCNCCNCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC 719
QY	ccngcgtgcggcgctgatgcctccgggaactggqccaacacaacagattcttctt 868
Db	
B718	CCCCCCCCCCCCCCCCNCCNCCCNCNCCCCCCCCCCCCCCCCCCCCNNCCNCC 659
QY	cgatcgttggtcaccaaccacctgacgcacctgtcaaacggcgcgcaagggttttc 928
Db	
B658	CCNCCNCCCCCCCCC 599
QY	gcggcgacacctgtctaccatggccgatlgatcgcgatgataccagcagccacact 988
Db	
B598	CCNCCNCCCCCCCCC 539
QY	actcccgcaagcgaggtagtaccaccaggcgtcttcgaccagcgacacccctggcg 1048
Db	
B538	CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCNCCCCCCC 479
QY	tgtatcacacaaacgcctagcctccccggcccgagggatcgtgttgcgaacgacc 1108
Db	
B478	CNCCC 419
QY	gcggctcacaaaaccggctgtgagcaccggcctcatacagaccgaagcccaacgtca 1168
Db	
B418	CCNCCNCCCCCCCCC 359
QY	ccgcctggacagcagccggagcagccagcatcacccagctgacccctgctctgcggcc 1228

[illegible]

[illegible]

Search completed: April 2, 2002, 21:47:49
Job time: 10814 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 22:47:50 ; Search time 366.19 Seconds
(without alignments)
1060.565 Million cell updates/sec

Title: US-09-785-904-2

Perfect score: 453

Sequence: 1 gatcgcgagcgacatca.....cgggctgcgcgcgtgatc 453

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_1101:*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	408.4	90.2	1535	16 AAQ89200 Mycobacterium tube
2	408.4	90.2	1535	17 AAT33656 M. tuberculosis ce
3	408.4	90.2	1535	19 AAV18647 DNA for M. tubercu
4	366.8	81.0	12412	17 AAT33537 BCG deletion regio
5	351	77.5	650	17 AAT33658 M. tuberculosis ma
6	351	77.5	650	19 AAV18649 DNA for M. tubercu
7	57.2	12.6	30690	21 AAQ92301 S. avermitilis ave
8	55.6	12.3	12381	21 AAZ58381 Streptomyces averm
9	55.6	12.3	31422	21 AAA92302 S. avermitilis ave
10	52.4	11.6	3198	20 AAX02974 Human li-lra BAC c
11	49.4	10.9	77536	21 AAA14651 Nucleotide sequenc

12	47.8	10.6	2697	13 AAQ22957 Sequence of beta-1
13	47.2	10.4	1000	21 AAA02484 Human colon cancer
14	47.2	10.4	1454	20 AAZ17038 Human gene express
15	46.6	10.3	1337	20 AAZ17263 Human gene express
16	45.8	10.1	114955	20 AAX53491 Human adenosine A1
17	45.6	10.1	4466	21 AAA14663 Nucleotide sequenc
18	45.6	10.1	4478	21 AAA14661 Nucleotide sequenc
19	45.6	10.1	4547	21 AAA14664 Nucleotide sequenc
20	45.6	10.1	4571	21 AAA14662 Nucleotide sequenc
21	45.6	10.1	77536	21 AAA14651 Nucleotide sequenc
22	45.4	10.0	2888	17 AAT59268 Streptomyces prist
23	45	9.9	567	21 AAA29550 HIV codon altered
24	44.6	9.8	6085	18 AAT70153 S.Lonqisporoflavus
25	43.4	9.6	1028	13 AAQ27091 XTY26 probe. Homo
26	43.4	9.6	1028	22 AAQ1542 1.0kb PstI fragmen
27	43.4	9.6	1028	22 AAS01347 Human Fragile X Sy
28	43.2	9.5	1107	22 AAH74540 Nucleotide sequenc
29	43.2	9.5	2712	20 AAZ08825 Streptomyces albid
30	43.2	9.5	2712	22 AAH74538 Nucleotide sequenc
31	42.6	9.4	1018	16 AAT02326 Human cardiac hype
32	42.6	9.4	1018	18 AAT87014 Human cardiotoxphi
33	42.6	9.4	1203	20 AAZ00869 Human c-Maf cDNA.
34	42.6	9.4	1485	21 AAZ61590 DNA encoding a hum
35	42.6	9.4	1485	21 AAZ61590 DNA encoding a hum
36	42.6	9.4	1539	20 AAX87267 cDNA clone encodin
37	42.6	9.4	1539	21 AAA99914 cDNA encoding huma
38	42.6	9.4	1539	21 AAA72705 Human cardiotoxphi
39	42.6	9.4	1539	21 AAA72706 Human cardiotoxphi
40	42.6	9.4	1539	21 AAA46936 cDNA encoding nove
41	42.6	9.4	1539	22 AAC90578 Human PRO882 cDNA.
42	42.4	9.4	494	19 AAV64542 M. tuberculosis im
43	42.4	9.4	494	19 AAV44433 Mycobacterium tube
44	42.4	9.4	494	20 AAZ19343 M. tuberculosis an
45	42.4	9.4	494	20 AAZ19131 M. tuberculosis re

ALIGNMENTS

RESULT	1
AAQ89200	
ID	AAQ89200 standard; DNA; 1535 BP.
XX	
AC	AAQ89200;
XX	
DT	09-OCT-1995 (first entry)
XX	
DE	Mycobacterium tuberculosis DNA sequence encoding mammalian cell entry
DE	protein.
XX	
KW	Vaccine; tuberculosis; ss.
XX	
OS	Mycobacterium tuberculosis.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..1535
FT	/*Lag= a
XX	
PN	WU9506726-A.
XX	
PD	09-MAR-1995.
XX	
PF	01-SEP-1994; 94WO-US09863.
XX	
PR	02-SEP-1993; 93US-0118442.
XX	
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Riley LW;
XX	
DR	WPI; 1995-115442/15.
DR	P-PSDB; AAR71931.
XX	

[illegible]

```
QY 418 cagcagggctgctgccgggctgcgcgctgctgac 453
|||||
Db 1246 cagcagggctgctgccgggctgcgcgctgctgac 1281

RESULT 3
AAV18647
ID AAV18647 standard; DNA; 1535 BP.
XX
AC AAV18647;
XX
DT 03-JUL-1998 (first entry)
XX
DE DNA for M. tuberculosis cellular uptake protein fragment.
XX
KW Cellular uptake protein; vaccine; infection; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 1..1534
FT /tag= a
FT /note= "stop codon not given"
XX
PN WO9805784-A1.
XX
PD 12-FEB-1998.
XX
PF 06-AUG-1997; 97WO-US13056.
XX
PR 07-AUG-1996; 96US-0689411.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Chong P, Riley LW;
XX
DR WPI; 1998-145620/13.
DR P-PSDB; A:W47541.
XX
XX Mycobacterium tuberculosis DNA - confers ability to enhance uptake of
PT therapeutic agents e.g. antibiotics, also useful in vaccines
XX
PS Disclosure; Pages 9-10; 82pp; English.
XX
CC The present sequence encodes a Mycobacterium tuberculosis cellular
CC uptake protein fragment, which confers on M. tuberculosis an
CC ability to enter mammalian cells and to survive within macrophages.
CC The protein can be used in a vaccine to prevent M. tuberculosis
CC infection, and provide for the uptake in cells of, e.g.
CC antibiotics, DNA fragments or anti-neoplastic agents. Antibodies
CC raised against it can be used to treat mammals already exposed to
CC M. tuberculosis, to induce a passive immunity and prevent disease
CC occurrence.
XX
SQ Sequence 1535 BP; 297 A; 544 C; 458 G; 236 T; 0 other;

Query Match 90.28; Score 408.4; DB 19; Length 1535;
Best Local Similarity 97.8%; Pred. No. 1.1e-78;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

QY 1 gatccgacgagccacatcaagtgatcgcccttttgcaccctgcgcgcgcg 60
|||||
Db 827 gatccgacgagccacgctcaagtgatcgcccttttgcaccctgcgcgcgcg 886
|||||
QY 61 tggatgtgtccaccgagccgagccgagccgacctgcgcgcgcgcgcgcgcgc 118
|||||
Db 887 tggatgtgtccaccgagccgagccgagccgacctgcgcgcgcgcgcgcgcgc 946
|||||
QY 119 gtcccgacgagctgcccgcctacccagcggtcatgactgactgactacacccgacg 178
|||||
```



```
/*tag= a
/note= "stop codon not given"
```

FT		/**tag= a	
FT		/note= "stop codon not given"	
XX	XX		
XX	PX	WO9805784-A1.	
XX	PD	12-FEB-1998.	
XX	XX		
XX	PF	06-AUG-1997; 97WO-US13056.	
XX	XX		
XX	PR	07-AUG-1996; 96US-0689411.	
XX	XX	(CONN-) CONNAUGHT LAB LTD.	
XX	PA	(CORR) CORNELL RES FOUND INC.	
XX	PA		
PI	Chong P, Riley LW;		
XX	DR	WPI; 1998-145620/13.	
XX	DR	P-PSDE; AAW47543.	
XX	XX		
PT	Myco bacterium tuberculosis DNA - confers ability to enhance uptake of		
PT	therapeutic agents e.g. antibiotics, also useful in vaccines		
XX			
XX	Disclosure; Page 14; 82pp; English.		
PS			
XX			
CC	The present sequence encodes a Mycobacterium tuberculosis cellular		
CC	uptake protein fragment, which confers on M. tuberculosis an		
CC	ability to survive within macrophages.		
CC	The protein can be used in a vaccine to prevent M. tuberculosis		
CC	infection, and provide for the uptake in cells of, e.g.		
CC	antibiotics, DNA fragments or anti-neoplastic agents. Antibodies		
CC	raised against it can be used to treat mammals already exposed to		
CC	M. tuberculosis, to induce a passive immunity and prevent disease		
CC	occurrence.		
XX			
XX	Sequence 650 BP; 123 A; 258 C; 185 G; 84 T; 0 other;		
SQ			

XX	AAA92301;
AC	
XX	
XX	10-JAN-2001 (first entry)
DT	
XX	
XX	S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
DE	
XX	
KW	Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW	multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW	agrochemical; ds.
XX	
XX	Streptomyces avermitilis.
OS	
XX	
XX	
FH	Key Location/Qualifiers
FT	CDS 1..11919
FT	/*tag= a
FT	/note= "avermectin aglycon synthase protein"
FT	CDS 11971..30690
FT	/*tag= b
FT	/note= "avermectin aglycon synthase protein"
ET	
XX	
PN	WO2000050605-A1.
XX	
XX	31-AUG-2000.
PD	
PD	
XX	23-FEB-2000; 2000WQ-IP01041.
PF	
XX	
PR	24-FEB-1999; 99JP-0046961.
XX	
PA	(KITA) KITASATO INST.
XX	
PI	Omura S, Ikeda H;
XX	
XX	WPI; 2000-565458/52.
DR	
DR	P-PSDB; AAB23749, AAB23750.
XX	
XX	Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT	it for the production of avermectin and its derivatives for drug and
PT	agrochemical use
XX	
PS	Claim 2; Page 66-134; 314pp; Japanese.
XX	
CC	The present sequence represents DNA which encodes avermectin aglycon
CC	synthase proteins. Also described are: (1) polypeptides encoded by all
CC	or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC	cells transformed by the vectors; (4) preparation of the polypeptides
CC	by culture of the transformants; (5) preparation of avermectin aglycon
CC	or its derivatives by culture of transformed avermectin-producing
CC	microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC	containing sense or antisense sequences from the avermectin aglycon
CC	synthase DNA. The enzymes are useful for the production of modified
CC	forms of avermectin and of the intermediates in its biosynthesis, for
CC	use as drugs, veterinary drugs and agrochemicals.
XX	
SQ	Sequence 30690 HP: 5456 A; 12454 C; 8617 G; 4263 T; 0 other;

Query Match	12.6%	Score	57.2	DB	21	Length	30690
Best local Similarity	52.5%	Pred. No.	0.00083				

Qy	165	aacccgcagcggaacctcaccgacacgaaacggcgccgcaaacggcgcatcacccctgagg	224
Db	21951	acactcctcggcggaatcacgcgcgccaccccttcgcgggatcctcaacctcacgsgcgc	22010
Qy	225	aacacacaaacagcagcgcatgtcacggcctaagtggtacctgaccccccaagtctcggggc	284
Db	22011	caccacctcatcaccacaacggccaacctcatgcaaacattgcgcccgccagcaccttgc	22070

[illegible]

Query Match	77.5%	Score 351	DB 19	Length 650
Best Local Similarity	97.7%	Pred. NO. 1.9e-66		
Matches 388: Conservative	0	Mismatches 5	Indels 4	Gaps 3

Qy	60	gtgagtgtccaccgcgcaggccgcgaagccgacctgcgcgaaggc--ctcaatat	117
Db	1	gtgagtgtccaccgcgcaggccgcgaagccgacctgcgcgaaggcgcgtcaatat	60
Qy	118	cgccccagacgtgccgcgtacgccacgcgggtcatggacttgctaacccccagac	177
Db	61	cgccccagacgtgccgcgtacgccacgcgggtcatggacttgctaacccccagac	120
Qy	178	gacctcacgcacacccaagcgcgcgcgaacgc--gcattaccttgacaaacagctatgc	236
Db	121	gacctcacgcacacccaagcgcgcgcgaacgcggtacaccttgacaaacagctatgc	180
Qy	237	gacggatgtcacggttaagtgtacctgaccccccaagtgcgggcacatttgagac	296
Db	181	gacggatgtcacggttaagtgtacctgaccccccaag--cgcgggcacatttgagac	239
Qy	297	cgctctagccaactggcgcgcccgccgcgacacacccgcacgacacccccgtat	356
Db	240	cgctctagccaactggcgcgcccgccgcgacacacccgcacgacacccccgtat	299
Qy	357	cgacacacccccgatcgggcgccatcgacgcgcacacccgcagccaaqcccaacgcaa	416
Db	300	cgacacacccccgatcgggcgccatcgacgcgcacacccgcagccaaqcccaacgcaa	359
Qy	417	ccacacgggctgctggccgggctgcgcgcgtgctc	453
Db	360	ccacacgggctgctggccgggctgcgcgcgtgctc	396

RESULT 7
AAA92301
ID AAA92301 standard; DNA; 30690 BP.

FT /note= "f

FT	CDS	complement (9122..9883)
FT		/tag= h
FT		/note= "fkbh gene"
FT	CDS	complement (9894..10994)
FT		/tag= i
FT		/note= "fkbi gene"
FT	CDS	complement (10987..11247)
FT		/tag= j
FT		/note= "fkbj gene"
FT	CDS	complement (11244..12092)
FT		/tag= k
FT		/note= "fkdk gene"
FT	CDS	complement (12113..13150)
FT		/tag= l
FT		/note= "fkbl gene"
FT	CDS	complement (13212..23988)
FT		/tag= m
FT		/note= "fkbc gene"
FT	CDS	complement (13452..13662)
FT		/tag= n
FT		/note= "ACPe"
FT	misc_feature	complement (13761..14394)
FT		/tag= o
FT		/note= "KR6"
FT	misc_feature	complement (14517..15294)
FT		/tag= p
FT		/note= "ER6"
FT	misc_feature	complement (15438..16587)
FT		/tag= q
FT		/note= "dehydratase domain (DH) 6"
FT	misc_feature	complement (16587..17820)
FT		/tag= r
FT		/note= "acylttransferase domain (AT) 6"
FT	misc_feature	complement (17820..19053)
FT		/tag= s
FT		/note= "KS6"
FT	misc_feature	complement (19116..19326)
FT		/tag= t
FT		/note= "ACp5"
FT	misc_feature	complement (19464..20097)
FT		/tag= u
FT		/note= "KR5"
FT	misc_feature	complement (20241..21420)
FT		/tag= v
FT		/note= "DH5"
FT	misc_feature	complement (21420..22653)
FT		/tag= w
FT		/note= "AT5"
FT	misc_feature	complement (22653..23892)
FT		/tag= x
FT		/note= "KS5"
FT	CDS	complement (23992..46573)
FT		/tag= y
FT		/note= "fkbB gene"
FT	misc_feature	complement (24163..24373)
FT		/tag= z
FT		/note= "ACp4"
FT	misc_feature	complement (24997..26146)
FT		/tag= aa
FT		/note= "DH4 (inactive)"
FT	misc_feature	complement (26146..27430)
FT		/tag= ab
FT		/note= "AT4"
FT	misc_feature	complement (27430..28684)
FT		/tag= ac
FT		/note= "KS4"
FT	misc_feature	complement (28750..28960)
FT		/tag= ad
FT		/note= "ACp3"
FT	misc_feature	complement (29092..29740)
FT		/tag= ae
FT		/note= "KR3"
FT	misc_feature	complement (29869..31018)
FT		/tag= af
FT		/note= "DH3 (inactive)"
FT	misc_feature	complement (31018..32185)
FT		/tag= ag
FT		/note= "AT3"
FT	misc_feature	complement (32185..33439)
FT		/tag= ah
FT		/note= "KS3"
FT	misc_feature	complement (33505..33715)
FT		/tag= ai
FT		/note= "ACp2"
FT	misc_feature	complement (33823..34480)
FT		/tag= aj
FT		/note= "KR2"
FT	misc_feature	complement (34606..35749)
FT		/tag= ak
FT	misc_feature	complement (28750..28960)
FT		/tag= al
FT		/note= "DH2 (inactive)"
FT	misc_feature	complement (35749..37144)
FT		/tag= am
FT		/note= "AT2"
FT	misc_feature	complement (37145..38296)
FT		/tag= an
FT		/note= "KS2"
FT	misc_feature	complement (38371..38581)
FT		/tag= ao
FT		/note= "ACp1"
FT	misc_feature	complement (38677..39307)
FT		/tag= ap
FT		/note= "KR1"
FT	misc_feature	complement (39442..40609)
FT		/tag= aq
FT		/note= "DH1"
FT	misc_feature	complement (40609..41842)
FT		/tag= ar
FT		/note= "AT1"
FT	misc_feature	complement (41842..43093)
FT		/tag= as
FT		/note= "KS of extender module 1 (KS1)"
FT	misc_feature	complement (43144..43660)
FT		/tag= at
FT		/note= "ACP of loading domain"
FT	misc_feature	complement (43777..44629)
FT		/tag= au
FT		/note= "EK of loading domain"
FT	misc_feature	complement (44974..46573)
FT		/tag= av
FT		/note= "COA ligase of loading domain"
FT	CDS	46754..47784
FT		/tag= av
FT		/note= "fkbA gene"
FT	CDS	47785..52272
FT		/tag= ax
FT		/note= "fkbp gene"
FT	CDS	52275..71465
FT		/tag= ay
FT		/note= "fkba gene"
FT	misc_feature	52362..53576
FT		/tag= az
FT		/note= "KS7"
FT	misc_feature	53577..54716
FT		/tag= bd
FT		/note= "AT7"
FT	misc_feature	54717..55871
FT		/tag= ba
FT		/note= "DH7"
FT	misc_feature	56019..56819
FT		/tag= bc
FT		/note= "ER7"
FT	misc_feature	56943..57575
FT		/tag= bd
FT		/note= "KR7"

```
FT misc_feature 57710..57920 /tag= be
FT /note= "ACP7"
FT misc_feature 5790..59243 /tag= bf
FT /note= "KS8"
FT misc_feature 59244..60398 /tag= bg
FT /note= "AT8"
FT misc_feature 60399..61412 /tag= bh
FT /note= "DH8 (inactive)"
```

Query Match 10.9%; Score 49.4; DB 21: Length 77536;
Best Local Similarity 48.7%; Pred. No. 0.04;
Matches 134: Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
OY 135 ccgcctacgacgagcggtgatgactggtacaccccgagcgacactacgacacga 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39383 cgcgcctacgacgacacacacccctccacgacccacccctccacacacaccc 39324
OY 195 acgcgcgcgacgaacgcgacacacccctgacgaacacgacgaacgacgacga 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39323 acccagcagacacaccccccctccacacacgacgacacacacacacacac 39264
OY 255 actgactacacgaccccccacgacgacgacacacacacacacacacacacac 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39263 cctccgcgcgacacacacacacacacacacacacacacacacacacacac 39204
OY 315 cgcgcgcgcgacgacgaacacacacacacacacacacacacacacacacacac 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39203 cccacaccccccgcgacacacacacacacacacacacacacacacacacac 39144
OY 375 ggcgcgcacgacgacgacacacacacacacacacacacacacacacacacac 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39143 actgcgacacacacacacacacacacacacacacacacacacacacacac 39109
```

RESULT 12
AAQ22957
ID AAQ22957 standard; DNA: 2697 BP.
XX
AC AAQ22957;
XX
DT 24-JUL-1992 (first entry)
XX
DE Sequence of beta-1, 3-glucanase gene and flanking regions.
XX
KW Recombinant enzyme; glucanase; pUV5-GIS; ss.
XX
OS Oerskovia xanthineolytica.
XX
FH Key location/Qualifiers
FT RBS 448..452
FT /tag= a
FT sig_peptide 463..570
FT /tag= b
FT mat_peptide 571..2109
FT /tag= c
FT repeat_unit 2140..2155
FT /tag= d
FT /rpt_type= INVERTED
FT repeat_unit 2169..2184
FT /tag= e
FT /rpt_type= INVERTED

XX
PN W09203557-A.
XX
XX 05-MAR-1992.
XX
XX 16-AUG-1991; 91MO-CA00294.
XX
XX 17-AUG-1990; 90US-0568869.
PR

XX
PA (CANADA) NAT RES COUNCIL CAN.
XX
PI Shen SH, Chretien P, Bastien L, Sillaty SN;

XX
DR WPI: 1992-096900/12.
XX
DR P-PSDB: MAR22197.

XX
PT Recombinant beta-1,3-glucanase enzyme - produced by transforming
PT E. coli cells with expression vectors and free of protease
PT activity

XX
PS Example: Fig 2; 32pp; English.

XX
CC The inventors claim a recombinant enzyme prepn. comprising beta-1,
CC 3-glucanase or a mutant or variant of a vector contg. DNA encoding
CC it under the control of an exogenous promoter pref. lac UV5. The
CC example uses the native beta-1, 3-glucanase gene isolated from
XX Oerskovia xanthineolytica genomic DNA and expressed in pUV5-GIS.

XX
SQ Sequence 2697 BP; 402 A; 1004 C; 954 G; 337 T; 0 other;

Query Match 10.6%; Score 47.8; DB 13: Length 2697;
Best Local Similarity 46.5%; Pred. No. 0.075;
Matches 154: Conservative 0; Mismatches 177; Indels 0; Gaps 0;

```
OY 119 gtccgcacgacgtgacgacacacacacacacacacacacacacacacacac 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1409 ggcgcgcacgacacacacacacacacacacacacacacacacacacacac 1468
OY 179 acctacacgacacacacacacacacacacacacacacacacacacacacac 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1469 accacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1528
OY 239 cgcgcacgtacacacacacacacacacacacacacacacacacacacacac 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1529 agctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1588
OY 299 tgcctacacacacacacacacacacacacacacacacacacacacacacac 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1589 gcgcgcacgacgtgacacacacacacacacacacacacacacacacacac 1648
OY 359 acacacacacacacacacacacacacacacacacacacacacacacacacac 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1649 cgttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1708
OY 419 acgacgacgtgacgacacacacacacacacacacacacacacacacacac 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1709 ccgcgcacgtgacgacacacacacacacacacacacacacacacacacac 1739
```

RESULT 13
AAA02484/C
ID AAA02484 standard; cDNA: 1000 BP.
XX
AC AAA02484;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2475.
XX
DE Human: colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.

XX
OS Homo sapiens.

XX
XX W09958675-A2.
XX
XX 18-NOV-1999.
XX
PD

Query Match	10.4%	Score 47.2	DB 20	Length 1454
Best Local Similarity	40.4%	Pred. No. 0.099		
Matches 153	Conservative 0	Mismatches 225	Indels 1	Gaps 1
y	71	caccgcgcaggcgccgcgaagcgacactgcggcgaaaggcctcaatcgtgccgcgcagc	130	
b	1420	CCCNCGNCGGCCGCACACCCACACACACCCNACCCCTCCNCCCGCCCCCNCGTC	1361	
y	131	tggccgcgtacccgcagcgggtcatgtgactggctacacccccgcagcggcactcacgcaca	190	
b	1360	GNCACTCCNNNNNCNCCGCCACACNGNCCCTATCCCTCCCTCCGCCACACCNACGCGCN	1301	
y	191	cgcaacgcgcgcgcaaaacgcctcacctcgtgcaacagcaatacgcagcgcatgtgcacg	250	
b	1300	CCCCGNCCCCGCACANNCCCNCCNCCCCCCNCCACNCCCNCAACCAACCAACCCCC	1241	
y	251	gctaagtggctactgcaccccccaagtcgcgggcgccacctttgaagcgtgtgtagcacaac	310	
b	1240	CACNTNCGCCNCTCGCNCGCCNCCNNCCNCCGCCACGCANACNGCATCNCACCTCC	1181	
y	311	tggcgccccgcggcgacacacccccgcagcacacac-cgggtcatcgacacaccccc	369	
b	1180	GCCTGCCCCCCACAGNGACACGCCCNCCNCCNCAACACCCNGCTCNGCCTNCACG	1121	
y	370	gatcgggcgcatcgacccgcagcacccccgcagcgcgaagcccaacgcagcagcggtg	429	
b	1120	ACACCCACACCCCGCACCCNCCNCCNCCCATCTCTCGNCACCTCTNNNNCCNACACGC	1061	
y	430	ctggcgccggctggcggcgc	448	
b	1060	CCCNCAACNGCGCGCACNC	1042	

RESULT	15	
AAZ17263		
D	AAZ17263 standard; cDNA; 1337 BP.	
X		
C		
X	AAZ17263;	
X		
T	12-OCT-1999 (first entry)	
X		
E	Human gene expression product cDNA sequence SEQ ID NO:4735.	
X		
E	Human; gene; gene expression product; diagnosis; therapy; probe;	
W	detection; mapping; tissue typing; profiling; forensic; cancer;	
W	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.	
X		
X	Homo sapiens.	
X		
X	WO9938972-A2.	
N		
X		
D	05-AUG-1999.	
X		
F	28-JAN-1999; 99WO-US01619.	
X		
R	03-APR-1998; 98US-0080666.	
R	28-JAN-1998; 98US-0072910.	
R	24-FEB-1998; 98US-0075954.	
R	31-MAR-1998; 98US-0080114.	
R	03-APR-1998; 98US-0080515.	
X		
X	(CHIR) CHIRON CORP.	
A	(HYSE-) HYSEQ INC.	

X Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
I Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
I Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
I Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
I Stache-Grain B, Sudduth-Klinger J, Williams LT;
X WPI: 1999-494092/41.
X
X Novel human genes and their expression products which are
T

PT differentially expressed in different cell types
XX
PS Claim 1; Page 2250-2251; 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AA12532 to AA17779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA12532 to AA17779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

SQ Sequence 1337 bp; 42 A; 577 C; 27 G; 22 T; 669 other;

Query Match	10.38	Score	46.6	DB	20	Length	1337
Best Local Similarity	29.5%	Pred. No.	0.13				
Matches	124	Conservative	0	Mismatches	296	Indels	0
						Gaps	0

[illegible]

Search completed: April 2, 2002, 22:49:09
Job time: 9209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:46:33 ; Search time 165.4 Seconds
(without alignments)
620.281 Million cell updates/sec

Title: US-09-785-904-2
Perfect score: 453
Sequence: 1 gatcggagcgacacatca.....ccgggtgcgcgctgac 453

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	408.4	90.2	1535	US-08-464-052-1	Sequence 1, Appli
2	408.4	90.2	1535	US-08-461-002-1	Sequence 1, Appli
3	408.4	90.2	1535	US-08-689-411-1	Sequence 1, Appli
4	408.4	90.2	1535	PCT-US94-09863-1	Sequence 1, Appli
5	409.4	90.2	4411529	US-09-103-840A-1	Sequence 1, Appli
6	366.8	81.0	12412	US-08-390-878-18	Sequence 18, Appli
7	351	77.5	650	US-08-464-052-5	Sequence 5, Appli
8	351	77.5	650	US-08-461-002-5	Sequence 5, Appli
9	351	77.5	650	US-08-689-411-5	Sequence 5, Appli
10	52.4	11.6	152331	US-09-128-155-16	Sequence 16, Appli
11	49.4	10.9	4403765	US-09-103-840A-2	Sequence 2, Appli
12	48.6	10.7	4403765	US-09-103-840A-2	Sequence 4, Appli
13	44.6	9.8	6085	US-09-029-603-4	Sequence 4, Appli
14	43.6	9.6	4411529	US-09-103-840A-1	Sequence 1, Appli
15	43.4	9.6	1028	US-08-118-200-1	Sequence 1, Appli
16	43.4	9.6	1028	US-08-458-745-1	Sequence 1, Appli
17	43.2	9.5	2712	US-09-025-691-4	Sequence 4, Appli
18	42.6	9.4	1018	US-08-444-083-6	Sequence 6, Appli
19	42.6	9.4	1018	US-08-286-304-6	Sequence 6, Appli
20	42.6	9.4	1018	US-08-442-745-6	Sequence 6, Appli
21	42.6	9.4	1018	US-08-443-129-6	Sequence 6, Appli
22	42.6	9.4	1018	US-08-443-952-6	Sequence 6, Appli
23	42.6	9.4	1018	US-08-443-130-6	Sequence 6, Appli
24	42.6	9.4	1018	US-08-898-911-6	Sequence 6, Appli
25	42.6	9.4	1018	PCT-US93-04467-6	Sequence 6, Appli
26	42.6	9.4	1203	US-09-086-010-1	Sequence 1, Appli
27	41.8	9.2	1140	US-09-105-537-15	Sequence 15, Appli

28 41.8 9.2 3292 3 US-09-320-878-22 Sequence 22, Appli
c 29 41.8 9.2 3765 3 US-07-705-490-1 Sequence 1, Appli
c 30 41.8 9.2 3765 4 US-07-751-891B-1 Sequence 1, Appli
c 31 41.8 9.2 4362 2 US-08-455-073A-1 Sequence 3, Appli
c 32 41.8 9.2 13613 4 US-09-105-537-3 Sequence 138, App
c 33 40.6 9.0 882 4 US-08-818-112-138 Sequence 4, Appli
c 34 40.4 8.9 530 3 US-08-758-662-4 Sequence 6, Appli
c 35 40.4 8.9 897 4 US-09-434-288-6 Sequence 7, Appli
c 36 40.4 8.9 4524 2 US-08-845-998-7 Sequence 7, Appli
c 37 40.4 8.9 4524 3 US-09-206-537-7 Sequence 7, Appli
c 38 40.4 8.9 4524 4 US-09-430-854-7 Sequence 7, Appli
c 39 39.6 8.7 666 2 US-08-875-034A-1 Sequence 1, Appli
c 40 39.6 8.7 1294 4 US-09-312-038-4 Sequence 4, Appli
c 41 39.6 8.7 2289 4 US-09-312-038-3 Sequence 3, Appli
c 42 39.6 8.7 43280 2 US-08-804-227C-1 Sequence 1, Appli
c 43 39.4 8.7 2064 1 US-08-344-428-1 Sequence 1, Appli
c 44 39.4 8.7 11219 1 US-07-642-734C-1 Sequence 1, Appli
c 45 39.4 8.7 11219 3 US-08-439-009A-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-464-052-1
; Sequence 1, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; Uptake of Mycobacterium Tuberculosis and Uses thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 40,727
; REFERENCE/DATA KEY NUMBER: 19605/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-464-052-1

Query Match 90.2%; Score 408.4; DB 3; Length 1535;
Best Local Similarity 97.8%; Pred. No. 3.3e-80;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;
Qy 1 gatcggagcgacacatcaaaagtattcgccctttttccacactgcgcgcg 60
|||||
Db 827 GATCGCGGAGCGCACGCTCAAAGTGATTCGCGCCCTTTTTCGCCACCTGCGCGCGG 886

Qy	61	tggatgtgtccaccgcgagcgccgcgaagcagacctgccggcaaaaggc--ctcaatatc	118
Db	887	TGGATGTGTCCACCCGCCAGGCGCGCGAAGCGACTGGCGGCAAAAGCGCTCAATATC	946
Qy	119	gtcccagacagtggcccgctaagccagtcggtcatgacttgactcaccccaagcg	178
Db	947	GTCCCGACGAGCTGGCCCGCTACGCCAGCGGCTCATGGACTTGCTACACCCGACCGCG	1006
Qy	179	acctcacgcacacgaacgcgcgcgaacgc- gcatcacctctgagcaaccagcaatacg	237
Db	1007	ACCTCACGCACACGAACGCGGCCCAAAGCGGCGATCACCTTGAGCAACAG..AATACG	1066
Qy	238	ucggcatgtcacggtctaagtgtacttgcaccccccaagtgcggyggccaactttgaagcc	297
Db	1067	ACGGCATGTCA CGGCTTAAGTGGCTACTTGACCCCCCAAG -CGCGGGCGACTTTGAAGCC	1125
Qy	298	gtg .tagcacaactgcgcgcgcgcgcgcgaccaaccgcgacgacacaccccggtcac	357
Db	1126	GTGCTAGCCAATACTGGCGCGCCCGGGCGGACACACCOCGACGACACACCCCGGTGATC	1185
Qy	358	gacaccacccccgatggcgcgcattcgaccgcgacaccccgagcccaagcccacacgaac	417
Db	1186	GACACACCCCAGTGGCGCGCGCATCGAACCGGACACCCGACGCCAAGCCACGCAAC	1245
Qy	418	cacgaocgggctgtggccggggtgcgcgcgtgatc	453
Db	1246	CACGACGGGCTGCTGGCCGGGCTGGCGCGGTGATC	1281

RESULT 2
US-08-461-002-1
: Sequence 1, Application US/08461002
: Patent No. 6214543
: GENERAL INFORMATION:
: APPLICANT: Riley M.D., Lee W.
: TITLE OF INVENTION: DNA Molecule Encoding for Cellular
: TITLE OF INVENTION: Uptake of Mycobacterium Tuberculosis and Uses thereof
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle
: STREET: Clinton Square, P.O. Box 1051

```

STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,002
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/186 (D-1485B)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-461-002-1

```

Query Match 90.2%; Score 408.4; DB 4; Length 1535;
Best Local Similarity 97.8%; Pred. No. 3.3e-80;

Matches	446:	Conservative	0:	Mismatches	6:	Indels	4:	Gaps	3:
Oy	1	gatcggagcgccacatcaaaagtattcgccccttttgcgccacctgccccgcgcgg	60						
Dd	827	GATCGGAGGGCAGGTCAAAGTGATTGGGCCCTTTTTCGCCCACTGCCGC GCGCG	886						
Oy	6.	tggatgtgtccaccgcgcgaaggcgcgaagaagcacctgcgcggcaaaagc-	118						
Dd	887	TGGATGTGTCACCCGCGACGCGUCUGGAAGCGCACCTTGGCGGCAAAAGCGCTCAATAATC	946						
Oy	119	gtcccqarqaagcttgcccgctlacgcgcagcgggtcatlggacttgttacaccgcgcgcgc	178						
Dd	947	GTCCGACGACGTGGCCCCGCTACUCCACGCGGTGTCATGGACTGGCTTACCACCGCGACGCG	1006						
Oy	179	acctcaacqacaragaa'rgcgcgcgcgaacacgc-gcatcaccttgagcaaccagcatacag	237						
Dd	1007	ACCTCACGACACGAAAGCGGCGCGC:AAACCGCGCATACCTTGAGCAACACGCAATACG	1066						
Oy	238	acgcatagtcacgcttaagltggctacctgaccoccaaagtcgcggccacatttaaacc	297						
Dd	1067	ACGGCATGTCACGGCTAAGTGGCTACCTTGACCCCCCAG-CGCGGCCACCTTTGAAGCC	1125						
Oy	298	gtgttagccaactlpgcgcgcgcgcgcgcgcaccaaccocgacgaccacaccgcggtcatc	357						
Dd	1126	GTGTTAGCCAACA'TGGCGCGCGCGCGCGACCAACCGGACGACCACACCGCGGTCA TC	1185						
Oy	358	gacaccacccrriatgcgcgcgcacatcgaccgcgaaccgcgagccaaagccacaacgcac	417						
Dd	1186	GACACCA'CCCCATGCGGCGCGCATVGCACCGGACGACACCGCGGACGACGCGCAACGCCAAC	1245						
Oy	418	cacgaagggtctgtgcgcgggtctgcgcgctgac	453						
Dd	1246	CACGACGGGTGTGTGGCGGGGCTGCGCGCCTGATC	1281						

RESULT 3

```

US-08-689-411-1
; Sequence 1, Application US/08689411
; Patent No. 6224881
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; APPLICANT: Chong, Pete
; TITLE OF INVENTION: DNA MOLECULE FRAGMENTS ENCODING FOR
; TITLE OF INVENTION: CELLULAR UPTAKE OF MYCOBACTERIUM TUBERCULOSIS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,411
; FILING DATE:

```

CLASSIFICATION: 336
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/187
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SFO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535 base pairs
TYPE: nucleic acid


```
Best Local Similarity 97.8%; Pred. No. 6.6e-80;
Matches 446; Conservative 0; Mismatches 6; Indels 4; Gaps 3;

Qy 1 catcgccgagcgacatcaaaagtattcgccgcccctttttggcccaactgcccgcgcgg 60
Db 1789460 GATCGGCGAGCGCAGCGTCAAAAGTATTCGCGCCCTTTTTCGCGCCACCTGCCCGCGCGG 1789401

Qy 61 tggatgtgtccaccgcccagcgccgaagcgcgacgttgcgcgcaaaagc--ctcaaatatc 118
Db 1789400 TGGATGTGTCCACCGCGCAGCGCGCGGAAGCGACCTGCCCGCAAGCGCTCAATATC 1789341

Qy 119 gtcccgacgaactggccgctacccagcggttcattggaactggtactacaccccgagcg 178
Db 1789340 GTCCCGACGAGCTGGCCCGGTACGCGCCAGCGGTCTATGGACTGGCTACACCCCGACGCGG 1789281

Qy 179 acctcaccgacacgaagcgcgcgcgcaaacgc-gcatcacccctgagcaaacgaataacg 237
Db 1789280 ACCTCACCAGACCGGACCGCGCGCAAGCGCGGATCACCTTGACCAACCAATACG 1789221

Qy 238 acgcatgtcacgctgaagtgtactctgaccccccaagtcggtggccaccttttgaagcc 297
Db 1789220 ACGGCATGTACCGCTAAAGTGGCTACTGTACCGCCCAAG-CGCGGGGCCACCTTTTGAAGCC 1789162

Qy 298 gtgctagccaaatggccgcccgcgcgcgcaaacaccccgacgacacaccccggttcac 357
Db 1789161 GTGCTAGCCAAACTGGCCGCGCCCGCGCGGACCAACCCGACGACACACCGGTTCATC 1789102

Qy 358 gacacacccccgatgctggttcacctgaccccccaagtcggtggccaccttttgaagcc 417
Db 1789101 GACACACCCCGCATGCGCGCGCATCGACCGCGACCGCGAGCGCAAGCCCAACGCAAC 1789042

Qy 418 cagcaggggtgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 453
Db 1789041 CACGACGGGTGCTGCGCGGGTGGCGCGGTGATC 1789006

RESULT 6
US-08-390-878-18
; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Hourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STR.FT: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 13371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12412 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-390-878-18

Query Match 81.0%; Score 366.8; DB 1; Length 12412;
Best Local Similarity 95.6%; Pred. No. 3.7e-71;
Matches 432; Conservative 0; Mismatches 12; Indels 8; Gaps 5;

Qy 9 aggcgcacatcaaaagtattcgccgccccttttgcgccc--acctgcccgcgcggtgagatg 66
Db 462 AGCGCGACATCAAAAGTATTTCGCGCCCTTTTTCGCGCCAAACCTTGCCTCGCGCGCGGTGATG 521

Qy 67 tgtccc--acctgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 122
Db 522 TGTCCAAAGCGCGTACGCGCGCGCAAGCGCGACCTGGCGCGCAACCGCTCAAAATATTCGTTC 581

Qy 123 cgacgagctggccgctacgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 182
Db 582 CGACGAGCTGGCGCGCTATGCGCGCGGTGCTATGGACTGGCTACACCGCGCGCGCGGTGCT 641

Qy 183 caccgacacgaucgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 241
Db 642 CACCGACATCAAGCGCGCGCGCAACCGCGCATCACCTTGACCAACCGCATACGCG 701

Qy 242 catgtcacgctgaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 301
Db 702 CATGTCAACCGCTAAAGTGGGTACTGTACCGCGCGCGCAAG-CGCGGGCGCGCGCTTGAAGCG 760

Qy 302 tagcaaaatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 361
Db 761 TAGCCAAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 820

Qy 362 ccacccccgatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 421
Db 821 CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 880

Qy 422 acggtgctggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 453
Db 881 ACGGGCTGTGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912

RESULT 7
US-08-464-052-5
; Sequence 5, Application US/08464052
; Patent No. 6008201
; GENERAL INFORMATION:
; APPLICANT: Riley M.D., Lee W.
; TITLE OF INVENTION: DNA Molecule Encoding for Cellular
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,052
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/185 (D-1485B)
; TELECOMMUNICATION INFORMATION:
```



```

: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 19603/187
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 650 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: US-08-689-411-5

Query Match      77.5%; Score 351; DB 4; Length 650;
Best Local Similarity 97.7%; Pred. No. 7.2e-68;
Matches 388; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

QY 60 gtgatgtgtccaccgccaggccgcgaagcagcactgtccggcaaaagc--ctcaatat 117
Db 1 tTGGATGTGTCCACCCGCCAGCGCGCGAAGCCGACCTGGCGCGCAAGCCGCTCAATAT 60
QY 118 cgtcccgacgagctggcccgctacgcccgagcggtgcatggactggctacaccccgacggc 177
Db 61 CGTCCCGACGAGCTGGCCCGCTACGCCACGCGGTGCTATGGACTACACCCCGACGGC 120
QY 178 gacctaccgagacacggagacgcccgcgcaaaagc--gcatacccttgagcaaccgcaatac 236
Db 121 GACCTCACCGACACCGAAGCTGCCCGCAAGCGCGGCATACCCCTGAGCAACCGCAATAC 180
QY 237 gacggcattgtcacggctaaagtggctacctgaccccccaagtgcgcgggccacctttgaagc 296
Db 181 GACGGCATGTACGGCTAAGTGGCTACTGNACCCCCCAAG--CGCGGGCCCACTTTTCAAGC 239
QY 297 cgtgctagccaaactgaccgcccccgcgcgacgacaaaccccgagcaccacccccgggtcat 356
Db 240 CGTGCTAGCCAAATGTGCCGCCGCCCGCGCGGAGCAAGCCCGACGACACACCCCGGTGAT 299
QY 357 cgacaccacccccgatgcggcccatcgaccgcgacaccgagccaaagcccaacgcaacaa 416
Db 300 CGACACCAACCCCGATGCGGCCGCCCATCGACCGGACACCCCGACGCAAGCCCAACGCA 359
QY 417 ccacgacgggctgtagccggggctgcgcgctgac 453
Db 360 CCACGACGGCTGCTGGCGCGGCTGCGCGCTCAATC 396

```

```

RESULT 10
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match      11.6%; Score 52.4; DB 3: Length 152331;
Best Local Similarity 45.6%; Prod. No. 0.003;
Matches 182; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 29 cgcgcgcctttttgcacccacattgcgcgcgcgcgtggatgtctccaccgccgcgagcgcgcga 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21884 cccccccccccccccccccccccccccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 21943

QY 89 agcgcacctgcgcacaaggcctcaatatctccgcagagctgcccgcgtacgcgccagc 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21944 cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22003

QY 149 gggctatgactggctacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22004 cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22063

QY 209 gcgcacacccctgcacaacgcagcaatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 268
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22064 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22123

QY 269 ccccccgaattgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22124 ccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22183

QY 329 ccaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22184 cccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22243

QY 389 gcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22244 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 22282

```

```

RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert L. D.
; APPLICANT: WHITE, Owen K.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366, 200007-00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match.	10.9%	Score 49.4;	DB 4;	Length 4403765;
Best Local Similarity	98.0%	Pred. No. 0.018;		
Matches 447;	Conservative	0;	Mismatches 5;	Indels 4;
				Gaps 3;
QY	1	gategcgqdgcgacatcaaatgattcgcgcctcttttctgcgcacatcgccgcgcgcgq	60	
Db	1780338	GAATCGGGGAGGCGGCACATCAAACTGATTCGCGCCCTTTTTCGCCACCTGCCCGCCGCG	1780279	

QY 61 tggatgtgtccaccgcgagccgagccgacacctgcccggcaaaagc---ctcaatatc 118
|||||
Db 1780278 TGGATGTGTCCACCGCCGAGCGCCGCAAGCCGACCTGGCGGCAAGACGCTCAATATC 1780219
|||||
QY 119 gtcccgacgagtggtccgcgtacgcccagcgggtctatgactggtctacacccccgacgagc 178
|||||
Db 1780218 GTCCCGAGAGTGGCCCGCTAGCCCGAGCGGGTCAATGAGTGGCTACACCCCGACGGCG 1780159
|||||
QY 179 acctaccgacagcaagccgcccgcgaacgc-gcatcacctttagcaacacgaataacg 237
|||||
Db 1780158 ACCTCACGACACCAAGACCGCGCGCAACGCGGCATCACCCCTGAGCAACGACATACG 1780099
|||||
QY 238 acgcatgtcacggctaaagtgtacctgaccccccaagtgcgggcccacacttttgaagcc 297
|||||
Db 1780098 ACGCATGTCTACGGCTAAGTGGCTACTGTACCCCCCAAG-CGCGGGCCACTTTGAAGCC 1780040
|||||
QY 298 gtgctagccaactggtccgccccggcgaccccaagtcgagccacacccccgggtcatc 357
|||||
Db 1780039 GTGCTAGCCAACCTGGCGCGCGCGCGGCGGCGGACCAACCCCGACGACACACCCCGGTGATC 1779980
|||||
QY 358 gacacacccccgatgcggcgccgcatgacgacgacaccccgagcccaagcccaacgcaac 417
|||||
Db 1779979 GACACCAACCCCGATGCGCGCGCGCATGGACCGGACACCCGAGCCAGCCCAACGCAAC 1779920
|||||
QY 418 lacgacgggctgctggcgggctgcgagcgctgctgac 453
|||||
Db 1779919 CACGACGGGCTGTGCGCGGGCTGCGCGCGCTGATC 1779884
|||||

RESULT 12
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 10.7%; Score 48.6; DB 4; Length 4403765;
Best Local Similarity 97.1%; Pred. No. 0.027; Mismatches 0; Indels 4; Gaps 3;
Matches 443; Conservative 0; Mismatches 0; Indels 4; Gaps 3;
QY 1 gatcgcgagcgccacatcaagtgttcgccccttttgcacactgcccgcgcg 60
|||||
Db 3869789 gatcgcgagcgccagtcgaagtgttcgccccttttgcacactgcccgcgcg 3869848
|||||
QY 61 tggatgtgtccaccgcgagccgcaagccgacacctgcccggcaaaagc---ctcaatatc 118
|||||
Db 3869849 tggatgtgtccaccgcgagccgcaagccgacacctgcccggcaaaagcgtctcaatatc 3869908
|||||
QY 119 gtcccgacgagtggtccgcgtacgcccagcggtctatgactggtctacacccccgacgagc 178
|||||
Db 3869909 gtcccgacgagtggtccgcgtacgcccagcggtctatgactggtctacacccccgacgagc 3869968
|||||

QY 179 acctaccgacacccgaacgcccgcgaacgc-gcatcacctttagcaacacgaataacg 237
|||||
Db 3869969 acctaccgacacccgaacgcccgcgaacgcgtatcacctttagcaacacgaataacg 3870028
|||||
QY 238 acgcatgtcacggctaaagtgtacctgaccccccaagtgcgggcccacacttttgaagcc 297
|||||
Db 3870029 acggtatgtcacggctaaagtgtacctgaccccccaag-cgcgggcccacacttttgaagcc 3870087
|||||
QY 298 gtgctagccaactggtccgccccggcgaccccaagtcgagccacacccccgggtcatc 357
|||||
Db 3870088 gtgctagccaactggtccgccccggcgaccccaagtcgagccacacccccgggtcatc 3870147
|||||
QY 358 gacacacccccgatgcggcgccgcatgacgacgacaccccgagcccaagcccaacgcaac 417
|||||
Db 3870148 gacacacccccgatgcggcgccgcatgacgacgacaccccgagcccaagcccaacgcaac 3870207
|||||
QY 418 cacgacgggctgtggcggggctgcgagcgctgctgac 453
|||||
Db 3870208 cacgacgggctgtggcggggctgcgagcgctgctgac 3870243
|||||

RESULT 13
US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; CURRENT FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
US-09-029-603-4

Query Match 9.8%; Score 44.6; DB 4; Length 6085;
Best Local Similarity 51.2%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 249 cggctaaagtgtacctgaccccccaagtcgcgggcgaacctttgaagccgtgtaagccaa 308
|||||
Db 5298 cgacgaggtgatcacggtgtcccaacacccgcccgcacacgctgctgcctacgactccac 5357
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2002, 21:47:49 ; Search time 3835.7 Seconds
(without alignments)
1269.087 Million cell updates/sec

Title: US-09-785-904-2

Perfect score: 453

Sequence: 1 gatcgcgagcgacatca.....ccggcgtcgcgcgctgac 453

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estfun:*

2: em_esthum:*

3: em_estin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estov:*

8: em_estov:*

9: em_htc:*

10: gb_est1:*

11: gb_est2:*

12: gb_htc:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pin:*

18: em_gss_pro:*

19: em_gss_rod:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	61.2	13.5	1014	13	CNS0057B	AL057121 Drosophil
2	59.6	13.2	1101	13	CNS014PA	AL104440 Drosophil
3	58.2	12.8	576	11	BG786264	BG786264 SEAMC006
4	57.4	12.7	1101	13	CNS0153F	AL104949 Drosophil
5	56.8	12.5	520	11	BG241147	BG241147 OVI_38_G1
6	56.6	12.5	848	13	CNS021YE	AL220271 Tetraodon
7	56.4	12.5	925	13	CNS0091P	AL053013 Drosophil
8	56.4	12.5	1101	13	CNS017SY	AL108460 Drosophil
9	56.2	12.4	644	13	CNS060D94	AL393518 T7 end of
10	56	12.4	632	11	BG320266	BG320266 Zm03_03g0
11	56	12.4	812	10	AL537183	AL537183
12	56	12.4	921	13	AZ211117	AZ211117 SP_0159_B

C 13	55.8	12.3	892	13	AZ200928	AZ200928 SP_0134_B
14	55.6	12.3	786	13	AZ196321	AZ196321 SP_1031_B
C 15	55.4	12.2	566	13	CNS03JN0	AL247077 Tetraodon
16	55.4	12.2	776	13	CNS010RY	AL099352 Drosophil
17	55.4	12.2	848	13	CNS0118A	AL099940 Drosophil
18	55.4	12.2	908	13	CNS006B4	AL064031 Drosophil
19	55	12.1	839	13	CNS004NB	AL054280 Drosophil
20	55	12.1	1016	11	BG368625	BG368625 HVSME1002
C 21	54.8	12.1	878	11	BG786339	BG786339 SEAMC006
22	54.8	12.1	1038	13	CNS015VW	AL105973 Drosophil
23	54.4	12.0	631	11	BG413362	BG413362 7037d09_x
C 24	54.4	12.0	666	11	BG786337	BG786337 SEAMC006
C 25	54.2	12.0	745	13	AZ046300	AZ046300 ubeu0091L
C 26	53.8	11.9	892	13	AZ194728	AZ194728 SP_1028_B
C 27	53.6	11.8	899	11	BG076530	BG076530 H3001D05-
C 28	53.6	11.8	895	13	CNS0071A	AL066286 Drosophil
C 29	53.6	11.8	978	13	B09456	B09456 F2RA17-Sp6
C 30	53.6	11.8	999	10	AL525224	AL525224 AL525224
C 31	53.2	11.7	549	13	CNS04CNH	AL284678 Tetraodon
C 32	53	11.7	935	13	CNS006XK	AL066051 Drosophil
33	52.8	11.7	794	13	AZ185894	AZ185894 SP_1005_B
34	52.6	11.6	1028	13	CNS015AB	AL105197 Drosophil
C 35	52.4	11.6	673	13	AZ184272	AZ184272 SP_1002_B
C 36	52.4	11.6	741	13	AZ194927	AZ194927 SP_1029_A
37	52.4	11.6	775	13	AZ185516	AZ185516 SP_1005_A
38	52.4	11.6	795	13	AZ188474	AZ188474 SP_1012_A
C 39	52.4	11.6	802	13	CNS02R43	AL210108 Tetraodon
C 40	52.4	11.6	1016	13	CNS03JU9	AL249930 Tetraodon
C 41	52.4	11.6	1101	13	CNS0181P	AL108775 Drosophil
42	52.2	11.5	915	13	AZ195437	AZ195437 SP_1030_A
43	52.2	11.5	925	10	AL581447	AL581447
C 44	52.2	11.5	932	13	CNS0072Q	AL066742 Drosophil
C 45	52.2	11.5	1059	13	CNS0155J	AL105025 Drosophil

ALIGNMENTS

RESULT 1

CNS0057B

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR1K11 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL057121

VERSION

AL057121.1

KEYWORDS

GSS.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999)

Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammos in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

JOURNAL MEDLINE REFERENCE AUTHORS JOURNAL	yeast species for molecular evolution studies 20584711 3 (bases 1 to 644) Genoscope. Direct Submission Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvarum</i> , <i>Saccharomyces</i> <i>exiguus</i> , <i>Saccharomyces servazzii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces thermotolerans</i> , <i>Kluyveromyces</i> <i>lactis</i> var. <i>lactis</i> , <i>Kluyveromyces marxianus</i> var. <i>marxianus</i> , <i>Pichia</i> <i>angusta</i> , <i>Debaryomyces hansenii</i> var. <i>hansenii</i> , <i>Pichia sorbitophila</i> , <i>Candida tropicalis</i> and <i>Yarrowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
COMMENT	

```

FEATURES
source
Location/Qualifiers
1..644
/organism="Zygosaccharomyces rouxii"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0AA011D12"
/clone_lib="AR0AA"
/note="end : T7"
128 a 235 c 88 g 104 t 89 others
BASE COUNT
ORIGIN

```

[illegible]

RESULT 10
BG320266
LOCUS
DEFINITION
Zm03_03g01.A.Zm03_RAFC_ECOR_cold_stressed_maize_seedlings_Zea_mays
EST 27-FEB-2001
CDNA clone Zm03_03g01, mRNA sequence.
BG320266
BG320266.1 GI:13149944
VERSION
KEYWORDS
SOURCE
EST.
Zea mays.
Zea mays.
ORGANISM
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 632)
Simmonds,J.A., Singh,J.A., Piche,C., Cass,L., Couroux,P., De Moors
A., Harris,L.J., Hattori,J.I., Ouellet,T., Robert,L.S., Sprott,D.
and Tinker,N.A.

TITLE	Expressed Sequence Tags from Cold-Stressed Maize Seedlings Grown Under High Light Intensity
JOURNAL	Unpublished (2001)
COMMENT	Contact: Singh, J.A. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-Food Canada 960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada Tel: (613) 759-1662 Fax: (613) 759-1701 Email: singhja@em.agr.ca.
FEATURES	Location/Qualifiers 1. .632 /organism="Zea mays" /cultivar="CQ328" /db_xref="taxon:4577" /clone="Zm03.03g01" /clone_lib="Zm03-AAFC-ECORC_cold_stressed_maize_seedlings" /tissue_type="leaf, crown" /dev_stage="4-leaf" /note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: Eco RI; Site_2: Xho I; Corn seedlings at 4-leaf stage were exposed to low temperature/high light (100C/700-800uE/m2/s) for 4 days. plants were grown/treated by J. Simmonds/I. Cass. Library prepared by C. Piche using Stratagene kit."
BASE COUNT	102 a 326 c 85 g 67 t 52 others
ORIGIN	

[illegible]

RESULT 11	
AL537183/c	
LOCUS	
DEFINITION	13-FEB-2001
AL537183	812 bp mRNA
AL537183	LT1_FL013_FBrn1 Homo sapiens cDNA clone CS0DF019YL14 3
prime, mRNA sequence.	
ACCESSION	AL537183
VERSION	AL537183.1
KEYWORDS	GI:12800676
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)

COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 812
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF019YL14"
 /clone_lib="LAI_FLO13_FBRnl"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 82 a 102 c '14 g 69 t 45 others
 ORIGIN

Query Match 12.4%; Score 56; DB 10; Length 812;
 Best Local Similarity 43.3%; Pred. No. 0.14;
 Matches 182; Conservative 14; Mismatches 224; Indels 0; Gaps 0;

QY	29	cgcgcctttttgcgcacactgcgcgcgcggtggatgtgtccacccgcagcgcgcga 88
Db	511	CCA 452
QY	89	agcgcacctgcggaaagcctcaatatcgtccgcagcagtgccgcgtacgcacgc 148
Db	451	CCGCGCCCCCCCCCCCCCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392
QY	149	gggtcatggactggctacacccgcgcgcgcgcgcctccacgcacccgcgcgcgaac 208
Db	391	CCCCCCCMAGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 332
QY	209	gcgcatacctgagcaaccacgaatcacgcggcatgtcacggctgaagtggctacctgc 268
Db	331	MCACACCG 272
QY	269	cucccaagtgcgggcacactttgaagcgtgtagcacaactggcgcgcgcgcgcgcga 328
Db	271	CG 212
QY	329	cgaacccgcagcaccacaccccggttcacgacacaccccgatgcggtgcgtatgcacgc 388
Db	211	CCACCGCMWCCCCCG 152
QY	389	gcgcacccgcgcgcgaagcccaacgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 448
Db	151	CTCTCCCCCGMACCG 92

RESULT 12
 LOCUS A2211117/c
 DEFINITION SP_0159_B1_C05_SF6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-159 Col-9 Row=F, DNA sequence.
 A2211117
 KEYWORDS Strongylocentrotus purpuratus.
 SOURCE Strongylocentrotus purpuratus
 ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Echinoidea; Echinacea; Echinoida;

<hr/>						
ACCESSION	AZ200928					
VERSION	AZ200928.1	GI:8395946				
KEYWORDS	GSS.					
SOURCE	Strongylocentrotus purpuratus.					
ORGANISM	Strongylocentrotus purpuratus					
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;					
	Echinoidea; Euchinoidea; Echinacea; Echinoida;					
	Strongylocentrotidae; Strongylocentrotus.					
REFERENCE	1 (bases 1 to 892)					
AUTHORS	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.					
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources					
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)					
MEDLINE	20402566					
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 134 row: F column: 24 Seq primer: T7 Class: BAC ends High quality sequence stop: 892. Location/Qualifiers . . . 892 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="Plate=134 Col=24 Row=F" /clone.lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library." /note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli DH10B"					
FEATURES	source					
	BASE COUNT	47 a	175 c	542 g	119 t	9 others
	ORIGIN					
	Query Match	12.3%	Score 55.8:	DB 13:	Length 892;	
	Best Local Similarity	46.08;	Pred. No. 0.15;	Indels 0;	Gaps 0;	
	Matches 189;	Conservative 0;	Mismatches 222;			
QY	29	cgcgcccttttggccacactgcccgcggggtggatgtgtccaccgcagcgccga 88				
Db	778	CCCGCGGCCACCCACACCCCCTGGCCCCCCTCCGCCGCCGCCGCCGCCGCC 719				
QY	89	agcgcacctgcggcaaaaggcctaatatcgtccgcagcagagtggcgcctatcctc 148				
Db	718	ACCACCGGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCA 659				
QY	149	gggtcatggaactgctacaccgcagcggcacctcacgcagacccaagcgcgcgaac 208				
Db	658	CCGCCTCCCTCCCGCGAGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCA 599				
QY	209	gcgatcacctgagcaacagacaatacgaagcatgtaaggctaatggtacctgac 268				
Db	598	CCGCACCGGCCCGCGGCCCTCACCCCGCAGCCACCGCGCGCGCGCGCGCGCG 539				
QY	269	ccccaaagtgcggggccacattttaaggcgtgctagccaaactggcgcccccggcg 328				
Db	538	CCACACCGCCCGCGCGCCCGCCCGCACCGCGCGCGCCCGCGCGCGCGCGCGCA 479				
QY	329	ccaacccgcagaccacaccccggttcatagcacaccaccoccgatgcggcgcatg 388				
Db	478	CCACCTCCCTCCCGACAGCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 419				
QY	389	gcgacaccgcgagccaaagcccaacgcaaccagacggggtgctggcgggcg 439				
Db	418	CCGCGCGCGCCCGCGCGCCCGCCCGCGAGCCACCGCGCGCGCGCGCGCGCG 368				

